

CANADIAN JOURNAL OF PSYCHOLOGY

VOLUME 13, NO. 2

JUNE, 1959

Flicker: An unconditioned stimulus for imprinting: H. JAMES	59
Performance in a vigilance task with and without knowledge of results: P. D. McCORMACK	68
A note on the Hebb-Williams test of intelligence in the rat: GITA DAS & P. L. BROADHURST	72
Alcohol, alcoholism, and introversion-extraversion: MURIEL VOGEL	76
A non-parametric approach to the graphical analysis of trends: RICHARD H. WALTERS	84
A note on Lambert's "evaluational reactions to spoken languages": HENRI TAJFEL	86
Traits and college achievement: D. D. SMITH	93
Some ruminations of the validation of clinical procedures: PAUL E. MEEHL	102
Book reviews	127

PUBLISHED FOR THE
CANADIAN PSYCHOLOGICAL ASSOCIATION BY THE
UNIVERSITY OF TORONTO PRESS

AUTHORIZED AS SECOND-CLASS MAIL, POST-OFFICE DEPARTMENT, OTTAWA

Replacement
Direct
7-28-60
Shostak

CANADIAN JOURNAL OF PSYCHOLOGY

Editor: JULIAN BLACKBURN

Assistant Editor: P. H. R. JAMES

THE CANADIAN JOURNAL OF PSYCHOLOGY is published quarterly in March, June, September, and December. *Annual subscription*, \$4.00; *single number*, \$1.25.

Subscriptions. Orders and correspondence regarding subscriptions, change of address, and purchase of back numbers should be sent to:

THE SECRETARY-TREASURER, CANADIAN PSYCHOLOGICAL ASSOCIATION

Box 31, Postal Station D., Ottawa, Ontario

Contributions. Original manuscripts and correspondence on editorial matters should be sent to:

THE EDITOR, CANADIAN JOURNAL OF PSYCHOLOGY

Queen's University, Kingston, Ontario.

Information for contributors. The *Journal* publishes experimental and theoretical articles in all recognized fields of psychology. Contributors need not be members of the CPA or residents of Canada. Manuscripts should be submitted *in duplicate*, they should be double-spaced throughout, and they should follow standard practice as regards tables, references, etc. "Immediate publication" (i.e. in the next issue to go to press) can be arranged for authors willing to pay the extra costs involved.

CANADIAN PSYCHOLOGICAL ASSOCIATION, 1958-1959

President: DALBIR BINDRA, Montreal; *Past President:* W. E. BLATZ, Toronto; *President Elect:* G. H. TURNER, London; *Secretary-Treasurer:* F. R. WAKE, Ottawa; *Employment Bureau:* DR. H. DÖRKEN, 602 Jackson Bldg., Ottawa.

The Canadian Psychological Association also publishes *The Canadian Psychologist* which is distributed to members only. *Editor:* C. M. MOONEY, Box 121, Postal Station K, Toronto.

FLICKER: AN UNCONDITIONED STIMULUS FOR IMPRINTING¹

H. JAMES

Queen's University

THE YOUNG OF MANY BIRDS will follow the first moving object to which they are exposed during the early hours of their life, and will quickly form a preference for the company of this object (which may be almost anything from an ornithologist's hide (5) to a matchbox (3)) to that of others, including their natural parents. Lorenz (7) believed that this preference is acquired by a special process, which he called "imprinting," but other writers (2, 5) have taken the view that imprinting is not essentially different from ordinary conditioning. There is little evidence to support either of these assumptions since no direct comparisons cannot be made until we have some precise knowledge of the stimuli which elicit and control imprinting, and the experiments described here were made in an attempt to get this information. In particular I wished to test the hypothesis that an unconditioned stimulus (UCS) for imprinting is retinal flicker.

Several clues suggest that retinal flicker may be a critical factor, perhaps the most significant coming from some observations originally made by Menner (see Pumphrey, 8, pp. 185-186) on the functions of the pecten in the avian eye. The foliations of this vascular structure, which is roughly conical in shape and projects from its base at the blind spot towards the pupil, cast shadows on the retina, and Menner has shown that the presence of these shadows enhances the sensitivity of the eye to the movement of an image projected upon it. As the image moves in and out of the shadows, the level of illumination at the retina will rise and fall, and it is apparently to this fluctuation in illumination, rather than to any other aspect of the moving image, that the bird first responds. If this is the case, and if the hypothesis that flicker is a UCS for imprinting is correct, a flickering light should be as attractive to newly

¹The writer wishes to thank Dr. A. S. West, who extended to him the facilities of the Queen's Biological Station for the purpose of these experiments, and Miss Barbara Wiggin and Mr. H. Osser, who assisted in running the birds. The research was supported by grants from the National Research Council of Canada and the Committee on Scientific Research, Queen's University.

hatched chicks as a moving object is since the retinal effect of both will be the same. Furthermore, if flicker acts as a UCS, it should be possible to condition the chick to approach and possibly to follow an otherwise neutral object (CS) by consistently associating the latter with a flickering light source.

EXPERIMENT I

Apparatus

The chicks were housed in individual compartments in a brooder, each compartment measuring 6 in. \times 12 in. The brooder was continuously illuminated from above by 40-watt lights, one light to every 4 compartments, an arrangement which served to keep the chicks warm. Food and water were available at all times. The chicks were tested in a runway approximately 10 ft. long, 2 ft. wide, and 2 ft. 6 in. high, the floor of which was covered with sawdust, the walls lined with hardboard, and the top covered with a semi-transparent sheet of polythene. Four holes, $\frac{1}{2}$ in. in diameter, were drilled in a diamond pattern at each end of the runway, the diagonal distance between the holes being $4\frac{1}{2}$ in. and the centre of the diamond 5 in. from the floor of the runway. The holes were covered on the outside with semi-transparent polythene, and were illuminated from outside the runway by a pair of 7 $\frac{1}{2}$ -watt lights at each end. Timing relays were used to make the lights at one end of the runway flash continuously at one of the following on/off rates: 0.25/0.25 secs., 1.0/1.0 secs., or 5.0/5.0 secs. The lights at the other end of the runway were continuously lit. The 4 overhead lights in the room in which the chicks were run were lit throughout the experiment.

Subjects

Thirty-nine Barred Plymouth Rock chicks were obtained from a commercial hatchery and placed in the brooder overnight. They were given their first run in the apparatus the next morning, by which time they were approximately 48 hrs. old. Three days later a chick from one group was placed by mistake in the compartment occupied by a chick from another group, and the records of both had to be abandoned, since we had no way of distinguishing the chicks other than by the compartment they occupied. Six days after the start of the experiment a chick in the third group died. Hence the results reported here are for 36 chicks.

Procedure

The chicks were divided at random into 3 groups, each group being assigned to one of the flash rates mentioned above for the duration of the experiment. They were run individually, being given 2 trials a day for 7 days and 1 trial on the eighth day, with an interval of approximately 12 hrs. between trials. Each trial involved placing the chick in the centre of the runway, facing one of the 10 ft. side walls. Its distance from the end of the runway at which the light was flashing on and off was then measured to the nearest 3 in. every 30 secs. over a 5-min. period, at the end of which time it was removed from the runway and returned to the brooder. The chick's score for each trial was the mean of these distances. One of Gellerman's trial orders (4) was used to determine which end of the runway should be illuminated by the flashing light on any given trial.

RESULTS

The median distance of the chicks in each of the three groups from the end of the runway at which the light was flashing on and off is given for each trial in Figure 1. An analysis of variance using ranks (10, p. 184) of the total distance scores indicates that the differences between the

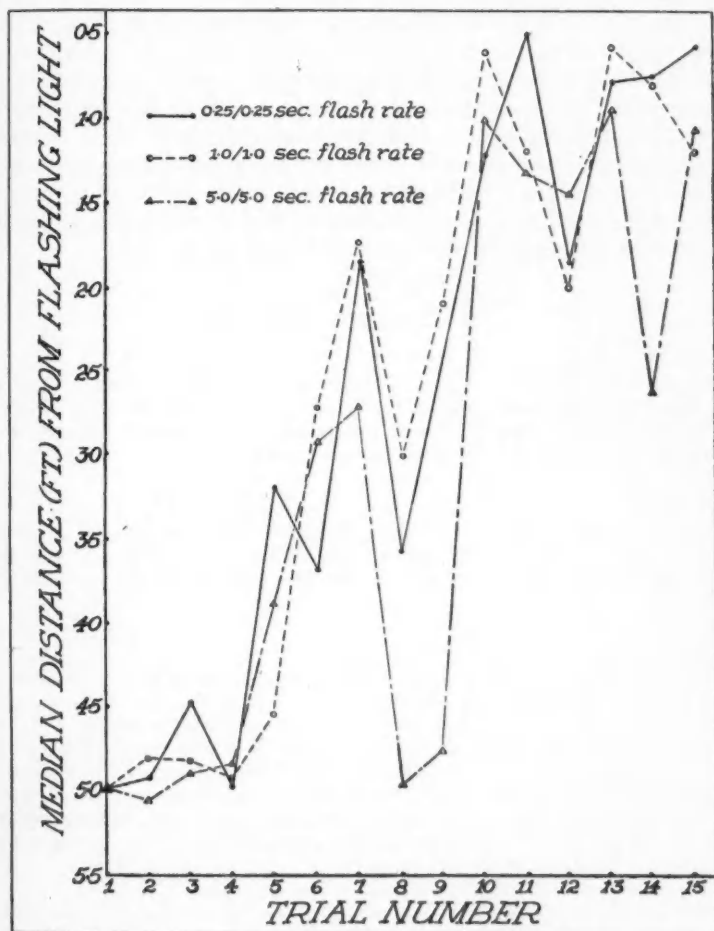


FIGURE 1. Median distance, in feet, of each group of chicks from that end of the runway at which the light was flashing, as a function of trial number (Experiment I).

groups are not significant ($\chi^2 = 0.5$). The main effect of the treatments seems to be shown in the distributions of the scores of each of the three groups, the Wald-Wolfowitz runs test (10, p. 136) giving a difference, significant at the 5 per cent level, for each of the three possible comparisons between the groups.

The behaviour of the chicks when they reached the end of the runway was sufficiently consistent to permit a general description. It took one of two forms: they either walked slowly up and down the end wall in the close vicinity of the holes, pecking at the sawdust on the floor; or else they stood still in front of the holes, pecking at the holes themselves or at the air just by them, as if they were trying to catch the light beam. Once or twice a chick was seen to settle down under the holes, ruffling out its feathers as if it were brooding. When the chicks came close to the end at which the light was flashing, they usually either fell silent or gave a soft "contentment" call, in marked contrast to the often piercing distress cries which they made as they were advancing up the runway.

EXPERIMENT II

Apparatus

The same apparatus was used as in Experiment I. On this occasion a turquoise polythene ball approximately 2½ in. in diameter was suspended by a nylon thread so that it hung 3½ in. from the floor of the runway. The ball could be moved along the length of the runway by pulling another nylon thread.

Subjects

Twenty Barred Plymouth Rock chicks, obtained from a commercial hatchery, were first run in the apparatus approximately 48 hrs. after hatching. One of the chicks died 3 days later, and the results reported here are for the remaining 19 chicks.

Procedure

The chicks were divided at random into an experimental ($N = 10$) and a control ($N = 9$) group. The chicks were run individually, and were given 2 five-min. trials a day for 5 days. For both groups, the light at one end of the runway flashed at an on/off rate of 0.25/0.25 secs., the light at the other end of the runway being steady. One of Gellerman's trial orders (4) was used to determine at which end of the runway the light should flash on and off on any particular trial. For the experimental group, the CS, a plastic ball, was always hanging against that end of the runway at which the light was flashing; for the control group the ball was hung against the end at which the light was steady. Otherwise, there was no difference in the treatment given to the two groups.

On the sixth day of the experiment, each chick was given 2 tests, half the chicks in each group taking the tests in one order and half in the other. Both tests were similar in that the light coming through the holes was now steady at *both* ends of the runway. They differed in the following respect. In one test, the ball remained

at the end of the runway throughout the 5-min. trial, and the distance of the chick from the ball was measured every 30 secs. In the other test, the ball was placed against one of the end walls and then, after 30 secs., was pulled silently 2 ft. down the runway at the rate of about 1 ft. every 2 secs. Thirty secs. later the ball was moved another 2 ft., and so on to the other end of the runway and back again. The distance of the chick from the ball was measured to the nearest 3 in. every 30 secs. In both tests, the chicks were started in the centre of the runway.

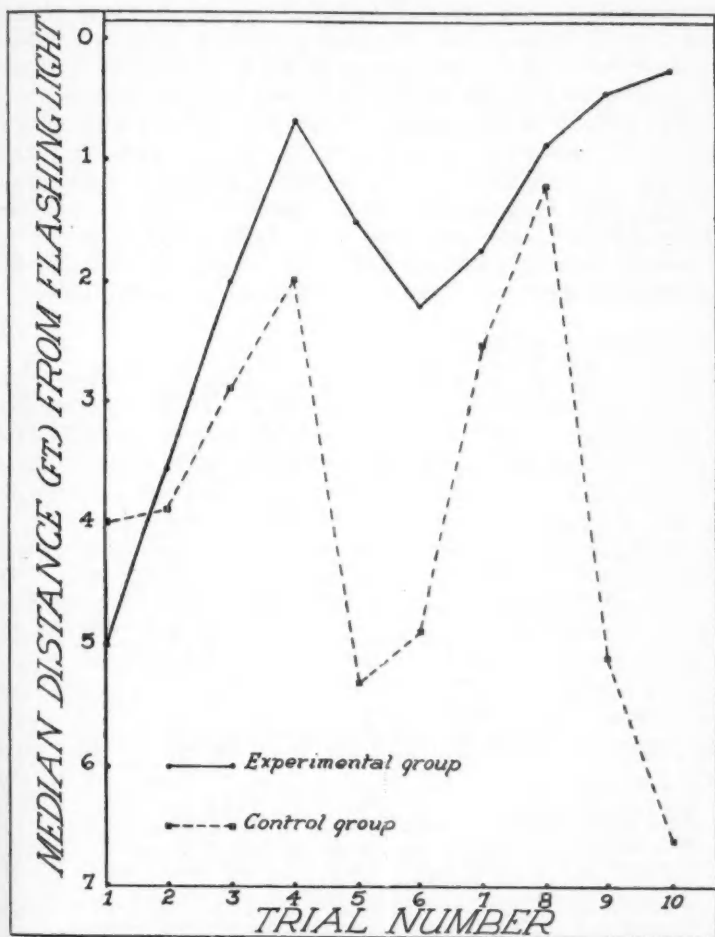


FIGURE 2. Median distance on each training trial of experimental and control group from that end of the runway at which the light was flashing (Experiment II).

RESULTS

Practice Trials (UCS and Stationary CS)

Figure 2 gives the median distance of the chicks from the flashing light for each of the ten practice trials. The difference between the total scores of the two groups is significant ($U = 18$; $0.05 > p > 0.02$, two-tailed test). Since the scores of the experimental group in this experiment (for whom the flashing light and the ball were contiguous) are slightly higher than the scores of the 0.25-second group in Experiment I (whose treatment was exactly comparable except for the absence of the ball), and since in the test with the stationary ball alone the control group preferred the end of the runway at which the ball was hanging (see below), it is reasonable to conclude that the ball itself had some properties which were attractive to the chicks. It is possible, therefore, that the control chicks in the present experiment were in a conflict situation, caught as they were between the flashing light at one end of the runway and the polythene ball at the other, and that this was responsible for their poorer performance during the practice trials.

Test with Stationary CS

The mean distance of each of the chicks from the stationary ball during the 5-minute test trial is given in Table I. The difference between the two groups is significant, on a one-tailed test, at the 0.025 level ($U = 19$). As their scores suggest, almost all the chicks spent the whole of the trial in that half of the runway where the ball was hanging; the only exceptions were two chicks (one in the experimental and one in the control group) who stayed in the centre of the runway and went to sleep and another control chick which ran to the opposite end of the runway during the last 30 seconds of the trial. There, however, the similarity between the behaviour of the two groups ends. Whereas the experimental chicks moved in the direction of the ball without retracing their steps towards the centre of the runway, pecked at the sawdust and made occasional contentment noises, the control chicks moved up and down the positive half of the runway giving intermittent distress cries.

TABLE I
MEAN DISTANCE (FT.) OF CHICKS FROM STATIONARY CS, EXPERIMENT II*

Chick	A	B	C	D	E	F	G	H	J	K
Experimental group	0.050	0.10	0.10	0.20	0.275	0.425	1.150	1.875	4.125	5.0
Control group	1.125	1.575	1.90	2.350	2.975	3.575	4.125	4.70	5.0	

*Distance between chicks and CS was measured to the nearest 0.25 ft. every 30 secs. over a 5-min. trial. Entries in table represent the means of the distances.

TABLE II

MEAN DISTANCE (FT.) OF CHICKS FROM MOVING CS, EXPERIMENT II*

Chick	A	B	C	D	E	F	G	H	J	K
Experimental group	1.0	0.150	0.525	1.275	0.025	0.750	0.475	2.825	2.40	2.850
Control group	3.575	4.10	2.725	4.750	4.925	1.750	2.60	3.90	2.50	

*Distance between chick and CS was measured to the nearest 0.25 ft. every 30 secs. over a 5-min. trial. Entries in table represent the means of these distances. Chicks have same letters assigned to them as in Table I.

TABLE III

RELATION BETWEEN MOVEMENTS OF CHICKS AND OF CS DURING SUCCESSIVE 30-SEC. PERIODS

	No. of times distance between chick and CS increased	No. of times distance between chick and CS decreased	No. of times chick stationary
Experimental group	16	65	9
Control group	33	26	22

Test with Moving CS

Table II gives the mean distance of each chick from the ball as it was moved up the runway and down again. The difference between the two groups is significant beyond the 0.01 level ($U = 9$) on a one-tailed test. Again the behaviour of the two groups was characteristically different. The experimental chicks either pecked at the sawdust in the immediate vicinity of the ball or pecked at the ball itself, and often came into bodily contact with it. The control animals, on the other hand, generally looked in the direction of the ball when it moved, but otherwise their movements appeared to be unrelated to those of the ball. This impression is supported by the data in Table III, which gives the relation between the movements of the ball and those of the chicks during successive 30-second periods of the trial. An example will make it clear how the entries in this table were computed. Suppose that at time t the ball was in the centre of the runway, and that 30 seconds later ($t + 1$) it moved two feet to the left. If at time t the chick was three feet to the right of the ball and at $t + 1$ had moved to the left two feet so that it was still three feet from the ball, it was considered to have decreased its distance from the CS (that is, it was nearer than it would have been if it had stayed still or moved in the opposite direction to that of the ball). It will be seen from Table III that while a high proportion of the experimental group's activity was such as to keep contact with the CS, the control group moved away from the ball about as often as they moved towards it.

DISCUSSION

The results of Experiment I indicate that flicker constitutes an adequate unconditioned stimulus² for approach behaviour, that the attractiveness of this stimulus increases with practice, and that the rate of flicker can be varied over a considerable range without appreciable effects. The results of Experiment II show that chicks will follow closely an object whose movements they would otherwise ignore if that object has, in the past, been consistently associated with a flickering light. This behaviour appears to be homologous with that of "imprinting," as the latter has been described by a number of writers (5, 6, 7).

These conclusions need to be qualified as follows. First, only Barred Rock chicks were used in the experiments reported in this paper. While we have, with the one exception noted below, been uniformly successful in getting chicks of this breed to approach and stay by a flickering light, we have been less successful in getting White Leghorn chicks to do the same, and quite unsuccessful with wild Mallard and Blue-winged Teal ducklings. Secondly, the shape, size, and brightness of the aperture through which the light was seen were not varied in the experiments described here. In an unpublished experiment, Barred Rock chicks, hatched and run at the same time and under similar conditions to those used in Experiment II, failed to show any consistent preference between a 60-watt lamp flashing with an on-off rate of 0.25/0.25 seconds behind a 6 in. diameter ground-glass window on which 1 in. wide vertical black stripes had been painted, and a similar aperture, without the stripes and illuminated by a steady light, at the other end of the runway.

Apart from the experiments which are implied by the cautionary statements made in the last paragraph, a number of other problems present themselves for study. The most obvious of these is the relation between the behaviour described here and that observed in the ordinary conditioning experiment. It cannot be too strongly emphasized that the homology between imprinting and classical conditioning has yet to be demonstrated. A second line of enquiry is suggested by some observations of Rheingold and Hess (9) on the relative attractiveness of mercury, plastic, water, and aluminum to White Rock chicks. Their conclusion that "attractiveness to the chick probably lies in a combination of a bright reflecting surface and the movement of the stimulus" is not far removed from our own, and suggests that a common physiological process

²It should be pointed out that in using the term "unconditioned stimulus" no reference is intended to the innateness of the behaviour elicited. I mean to imply no more than that (a) such a stimulus does not have to be paired with any other stimulus before it will elicit a predictable response, and that (b) the analogy with learning is worth pursuing experimentally.

may underlie approach to water and following behaviour. If this is so, it becomes an interesting question as to how the various responses which the chick makes to intermittent visual stimulation become differentiated. We do not recall seeing any of our chicks trying to drink the flashing light; yet Rheingold and Hess note that "it was instructive to observe this response (of drinking) to the metal and the plastic, where the beak would slide forward on the hard surfaces." Thirdly, the finding of Collias and Collias (1) that ducklings are attracted to a source of intermittent sound encourages the hope that this behaviour, artificially elicited by visual and auditory stimuli which can be precisely controlled and varied, will be a useful one in which to study the psychology and physiology of intersensory facilitation and transfer.

SUMMARY

Two experiments with newly hatched Barred Plymouth Rock chicks are described. In the first experiment it is shown that these chicks will approach an intermittent light source seen through 4 small holes at one end of a runway, and that the alacrity with which they do so increases with practice. In the second experiment it is shown that if a stationary conditioned stimulus is placed near this intermittent light source for 10 trials, the chicks will subsequently follow the conditioned stimulus up and down the runway. The results are taken to support the hypothesis that retinal flicker acts as an unconditioned stimulus for imprinting, a form of behaviour which appears to be homologous with that observed in these experiments.

REFERENCES

1. COLLIAS, N. E., & COLLIAS, E. C. Some mechanisms of family integration in ducks. *Auk*, 1956, 73, 378-400.
2. FABRICIUS, E. Some experiments on imprinting phenomena in ducks. *Proc. Xth Int. Ornith. Congr.*, 1951, 375-379.
3. FABRICIUS, E., & BOYD, H. Experiments on the following reaction of ducklings. *Wildfowl Trust Ann. Rep.*, 1952-53, 84-89.
4. HILGARD, E. R. Methods and procedures in the study of learning. Ch. 15 in S. S. STEVENS (ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
5. HINDE, R. A., THORPE, W. H., & VINCE, M. A. The following response of young coots and moorhens. *Behaviour*, 1956, 9, 214-242.
6. JAYNES, J. The interaction of learned and innate behavior: I. Development and generalization. *J. comp. physiol. Psychol.*, 1956, 49, 201-206.
7. LORENZ, K. Der Kumpan in der Umwelt des Vogels. *J. f. Ornith.*, 1935, 83, 137-213, 289-413.
8. PUMPHREY, R. J. The sense organs of birds. *Ibis*, 1948, 90, 171-199.
9. RHEINGOLD, H. L. & HESS, E. H. The chick's "preference" for some visual properties of water. *J. comp. physiol. Psychol.*, 1957, 50, 417-421.
10. SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. Toronto: McGraw-Hill, 1956.

PERFORMANCE IN A VIGILANCE TASK WITH AND WITHOUT KNOWLEDGE OF RESULTS^{1, 2}

P. D. McCORMACK

Defence Research Medical Laboratories, Toronto, Ontario

TWO STUDIES have been reported (1, 2) where the effects of knowledge of results on performance in a vigilance task have been investigated. In both of these, the mean number of missed signals and the increase, if any, in missed signal frequency over time were considerably less when the subject was informed of correct, missed, and false "detections" than when this information was withheld. In the light of certain criticisms which have been made (3) of the design of these two studies, it was decided to investigate performance in a vigilance task with and without knowledge of results in a situation more amenable to precise experimental design.

METHOD

Ten females were paid to serve as Ss. S's task was to depress a microswitch each time light from a 15-watt bulb was seen through an aperture placed at a distance of 12 ft. from her. Response time to the light was recorded by a Hunter Klockounter while the duration of the light (100 msec.) was controlled by a silent Hunter Decade Interval Timer. E was located in a partially sound-deadened cubicle, thus minimizing the presence of cues which might enable S to anticipate the onset of the light.

Before the experimental session began, S was instructed to keep her thumb on the switch and to depress it as fast as possible each time a light appeared. She was asked to do her best throughout the session which she was told would last for approximately one hour.

Following the instruction period, the light was presented 51 times to each of the 10 Ss, the intervals between stimuli being 30, 45, 60, 75, and 90 secs. The over-all inter-stimulus interval order was different for every S, with the restriction that all Ss experienced each of the 5 intervals once every 5 min. The interval sequence in every 5-min. block was selected at random from the 120 possible sequences.

S participated in the experiment on each of 2 consecutive days. Five of the Ss were provided with knowledge of results on the first day while the remaining 5 Ss received the knowledge condition on the second day. The order in which S received

¹Defence Research Medical Laboratories Report no. 234-4, DRML Project no. 234, PCC no. D77-94-20-42, HR no. 177.

²Appreciation is expressed to S/Sgt. E. A. Singer for processing the 10 Ss in the present study as well as for handling 60 male Ss in a pilot investigation involving between-subject comparisons, a design which subsequently proved to be untenable.

the 2 treatments was randomly determined. Under the knowledge-of-results condition, a red light flashed on a panel alongside the aperture each time *S* made a response which was slower than the preceding one. If a faster response was made, a green light appeared. Under the no-knowledge treatment, the red and green lights were not employed.

RESULTS AND DISCUSSION

The major findings of the investigation are summarized in Figures 1 and 2 where response time is plotted as a function of task duration and length of inter-stimulus interval for both the knowledge (K) and the no-knowledge (NK) conditions. The summary of an analysis of variance

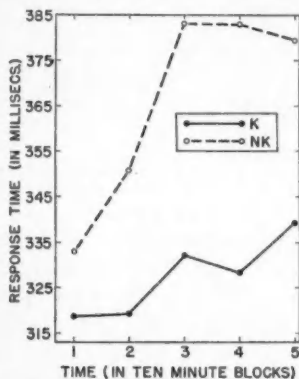


FIGURE 1. Response time as a function of successive 10-min. time blocks.

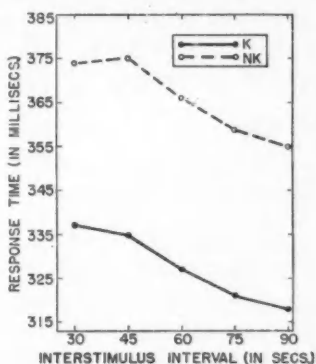


FIGURE 2. Response time as a function of length of inter-stimulus interval.

performed on the data is presented in Table I. (Note that the analysis is based on ten five-minute time blocks while these have been collapsed into five ten-minute blocks in Figure 1.) From an examination of this table, as well as of Figures 1 and 2, it is evident that the over-all level of performance under the knowledge-of-results condition was superior to that under the no-knowledge condition. It is also apparent that performance deteriorated as the session progressed, this effect being most obvious under the no-knowledge condition. Under both treatments, response time was found to be inversely related to length of inter-stimulus interval.

The subjects \times conditions interaction was somewhat inflated. However, this was almost entirely due to one subject. While all ten subjects responded faster under the knowledge than under the no-knowledge condition, this particular subject showed the effect in a more pro-

TABLE I
ANALYSIS OF VARIANCE OF RESPONSE TIME FOR ALL Ss ON EACH OF TWO CONSECUTIVE DAYS

Source	df	Sum of squares	Mean square
Subjects (S)	9	1,550,332.2	172,259.1*
Time blocks (T)	9	215,475.6	23,941.7*
Conditions (C)	1	369,139.4	369,139.4*
Inter-stimulus intervals (I)	4	59,660.2	14,915.0*
TC	9	92,872.5	10,319.2*
TI	36	117,905.6	3,275.2
CI	4	476.5	119.1
TCI	36	87,181.3	2,421.7
SC	9	—	—
S ₀ vs. rest × C	1	143,177.1	143,177.1*
Remainder	8	24,252.5	3,031.6
ST	81	383,684.7	4,736.8
SI	36	89,873.1	2,496.5
STC	81	335,649.0	4,143.8
STI	324	975,608.1	3,011.1
SCI	36	106,214.1	2,950.4
STCI (error)	324	902,965.1	2,786.9
TOTAL	999	5,454,467.0	

*Significant at the .001 level.

nounced fashion. The conditions \times intervals interaction was much smaller than would be expected on a chance basis. A significantly small effect such as this often indicates a lack of randomness somewhere in the experimental design. Since this was not the case in the present study, it is reasonable to assume that the significantly small CI interaction was obtained by chance. Although it is not shown in the analysis of variance summary table, mean response times were essentially the same on each of the two days and under each of the two orders in which the conditions were presented. The interaction effects of days or order with each of the remaining variables were treated as error.

There did not appear to be any marked signs of skewness in the response-time distributions nor was there any evidence for heterogeneity of variance. Also there were approximately an equal number of "slower" and "faster" responses at each of the five intervals and at each of the ten five-minute time blocks, which indicated that there was no serious confounding, under the knowledge condition, of the effects of these two types of knowledge with those of time blocks and intervals. It might also be mentioned that giving knowledge of results had a general rather than any specific effect on performance. This was reflected by the fact that the subject's mean response times following being told "slower" and "faster" did not differ significantly.

In an earlier study (3), where no knowledge of results was provided, performance deteriorated over time and improved following periods of

interpolated rest, but remained invariant with respect to length of inter-stimulus interval. These results were accounted for by postulating an inhibitory process which was assumed to develop continuously throughout the duration of the task, dissipating only during periods of interpolated rest. The findings of the present investigation do not seem to support this notion since, under both the knowledge and the no-knowledge conditions, a highly reliable inverse relation was demonstrated to hold between response time and interval length. Thus it is suggested that inhibition dissipates not only during periods of interpolated rest but between the presentation of stimuli as well, the rate of dissipation being the same whether knowledge of results is provided or not. On the other hand, inhibition appears to be generated at a faster rate under the no-knowledge than under the knowledge condition. A number of alternative interpretations of the data could be made at this stage; however, the one stated here appears to the investigator as the most promising working hypothesis.

SUMMARY

Ten females served as Ss in a vigilance task consisting of a 50-min. session on each of 2 consecutive days. Five of the Ss were provided with knowledge of results on the first day while the remaining 5 Ss received this knowledge on the second day. The order in which S received the 2 treatments was randomly determined. Under the knowledge-of-results condition, a red light was presented each time S made a response which was slower than the preceding one. If a faster response was made, a green light appeared. Under the no-knowledge treatment, the red and green lights were not employed. On both days, S was instructed to fixate on an aperture 1 cm. in diameter placed at a distance of 12 ft. from S and to depress a switch as fast as possible whenever light was seen through the aperture. The light was presented 51 times to each S each day, the intervals between stimuli being 30, 45, 60, 75, and 90 secs.

Response time increased significantly throughout the duration of the task, this increase being more pronounced under the no-knowledge than under the knowledge condition. Under both treatments, response time was found to be inversely related to length of inter-stimulus interval.

The findings of the present study were consistent with the hypothesis that inhibition is generated at a faster rate under the no-knowledge than under the knowledge condition and that the rate at which it dissipates between stimuli is the same regardless of whether or not knowledge of results is provided.

REFERENCES

1. BAKER, C. H. *Three minor studies of vigilance*. Defence Research Board of Canada, DRML Report no. 234-2, 1958.
2. MACKWORTH, N. H. *Researches on the measurement of human performance*. Med. Res. Council, Special Report no. 268, London: H.M. Stationery Office, 1950.
3. MCCORMACK, P. D. Performance in a vigilance task as a function of inter-stimulus interval and interpolated rest. *Canad. J. Psychol.*, 1958, 12, 242-246.

A NOTE ON THE HEBB-WILLIAMS TEST OF INTELLIGENCE IN THE RAT

GITA DAS¹ AND P. L. BROADHURST

Institute of Psychiatry (Maudsley Hospital), University of London

THE HEBB-WILLIAMS closed-field intelligence test for rats (2, 3) was used in an earlier study (1) in which the subjects were chosen from two strains selectively bred for high and low emotional defecation in the Hall open-field test (the Maudsley Reactive and Nonreactive Strains respectively). The intention was to establish whether or not the selection practised had inadvertently involved any differences in cognitive abilities also. The findings were negative. No differences attributable to sex or to degree of food deprivation were detected either. However, certain difficulties were encountered in the use of the Hebb-Williams test, particularly in attempting to use the assumed level of difficulty of the sub-tests as an experimental variable. It is the purpose of the present note to contribute to the standardization of the test by presenting data relating, first, to the level of difficulty of the twelve sub-tests or problems, and, secondly, to the order effects due to practice.

The first line of Table I gives the mean error scores for 46 albino rats (23 male and 23 female), aged 161.3 (\pm SD 11.5) days, from the selectively bred strains mentioned above who were given the test according to the standard procedure, including pre-training (1, 3), each under one of six different levels of hunger drive (1). From the figures given in the next line, it will be seen that the order empirically deter-

TABLE I
MEAN TOTAL ERROR SCORES FOR THE TWELVE PROBLEMS

	Problem												n
	1	2	3	4	5	6	7	8	9	10	11	12	
Over-all mean error score	4.3	9.3	5.7	7.0	20.2	12.6	17.8	13.9	16.5	16.6	16.7	16.0	46
Order of difficulty	1	4	2	3	12	5	11	6	8	9	10	7	
Mean error score (original order)	4.7	11.9	7.1	9.3	25.1	15.0	25.4	19.4	9.1	7.6	11.4	13.7	7*
Order of difficulty	1	7	2	5	11	9	12	10	4	3	6	8	

*n = 15 for problems 1-4 (see text).

¹Now at Utkal University, Orissa, India.

mined in this way departs from the assumed order quite markedly. But in order to distribute differences arising from possible order effects due to practice as evenly as possible among the sub-groups used in our study, three sets of four problems each were formed on the basis of the assumed order of difficulty and the six possible orders of these three sets distributed among our subjects. These sets comprised sub-tests 1 through 4 (A), 5 through 8 (B), and 9 through 12 (C). In order, therefore, to make a more precise comparison with the assumed order of difficulty, the scores of the seven subjects who were assigned the problems in the order of sets ABC were selected and are presented in the third line of Table I. The data for the first set of four problems (A) are based on an additional eight rats who were given the problems in the order ACB and whose scores for A (only) may, therefore, properly be included. As before, the resulting order of difficulty is given in the next line, and clearly differs from the one derived from the over-all scores.

This finding suggests the importance of order effects attributable to practice and requires further analysis. Table II gives the over-all mean total error score for each set of four problems in each position, and Table III a breakdown of these mean scores. There are six scores for each set, depending upon the position of the set in the series (first, second, or

TABLE II
MEAN TOTAL ERROR SCORE, ACCORDING TO POSITION, FOR
PROBLEMS GROUPED IN SETS OF FOUR

Position	Set			All
	A	B	C	
1	33.0	81.2	82.3	66.0
2	28.4	65.1	64.7	51.4
3	16.4	49.1	48.8	39.1
All	26.3	64.5	65.7	52.1

TABLE III
BREAKDOWN OF MEAN TOTAL ERROR SCORES FOR PROBLEMS
GROUPED IN SETS OF FOUR

Set A		Set B		Set C	
Order	Score	Order	Score	Order	Score
ABC	35.0	BCA	75.0	CAB	87.4
ACB	31.3	BAC	86.6	CBA	76.3
BAC	31.7	ABC	85.0	ACB	80.9
CAB	25.4	CBA	45.3	BCA	46.1
BCA	16.6	ACB	48.9	ABC	41.9
CBA	16.3	CAB	49.3	BAC	54.9

third), and which other set preceded (or followed) it. Two points emerge from these data. First, it will be seen from Table II that there are marked differences in the mean scores for a given set of problems depending on the position of the set in the order of administration, and, secondly, from Table III, that these differences are not dependent on the order alone, but rather on the sequence involved—that is, which set of problems preceded which. An analysis of variance confirms these impressions: effects attributable to position, to order, to sequence, and also to difficulty level of the sets, all yield F ratios significant at the .01 level or beyond. Thus, for set B, the mean score is consistently reduced to about half if set C precedes it, as compared with the value if set A precedes it, or if it is given in the first place. A similar trend is observed in set C, with reference to set B. In both cases a two-tail t test comparing the over-all mean of the first three orders shown with that of the second three yields a value significant beyond the .001 level. In Table II, it will be seen that these two sets (B and C) as formed are of comparable difficulty, whereas set A is much easier, so that it seems justifiable to say that a practice effect occurs with either set of difficult problems (sets B and C) but only as a result of practice on the other difficult problems, and not as a result of practice on the easier ones in set A.

A comparable effect of a single set of the more difficult problems upon the mean error score of the easier set A is not found. What is seen (Table III) is that when *both* sets of difficult problems precede the easier one, irrespective of the order in which these two preceding sets were administered, then there is a definite reduction in the error scores, and a two-tail t test comparing the over-all mean of the first four orders of set A with that of the last two shows that the difference is significant beyond the .02 level. This suggests that performance on the easier problems may be affected by practice, but that a larger amount of it is required to do so than is the case with the more difficult ones.

Two further points are worthy of mention. First, the reliability of the test as measured by the correlation between odd- and even-numbered problems was $+.48$, ($+.65$ when corrected for length by the Spearman-Brown formula). While this is somewhat lower than might have been expected from the test re-test reliability of $+.84$ found by Rabinovitch and Rosvold (3), it is not known how the variations in problem order introduced may have influenced this result. Secondly, it was noted that the correct paths from start to goal resemble each other closely in both problems 7 and 11, and problems 8 and 10, despite the differences in the positioning of the interposing barriers. It is considered that such similarity may have contributed to the order effects described above,

and it may be thought necessary to modify these particular problems in future use of the test. In any event, account must clearly be taken of the possibility of order effects due to practice in any such use involving predictions regarding the levels of difficulty of the constituent sub-tests.

REFERENCES

1. DAS, GITA, & BROADHURST, P. L. The effect of inherited differences in emotional reactivity on a measure of intelligence in the rat. *J. comp. physiol. Psychol.*, 1959 (in press).
2. HEBB, D. O. & WILLIAMS, K. A method of rating animal intelligence. *J. gen. Psychol.*, 1946, 34, 59-65.
3. RABINOVITCH, M. S. & ROSVOLD, H. E. A closed field intelligence test for rats. *Canad. J. Psychol.*, 1951, 5, 122-128.

ALCOHOL, ALCOHOLISM, AND INTROVERSION-EXTRAVERSION

MURIEL VOGEL¹

Alcoholism Research Foundation, Toronto

IN RECENT YEARS Eysenck (1, 2, 3) has attempted to obtain scientific evidence for the concept of an introversive-extraversive personality dimension, roughly similar to the concepts of Jung (12, 13) and McDougall (14). Eysenck has further attempted to relate his experimental findings to the existing body of psychological learning theory by means of the concept of "cortical excitation and inhibition." His use of the terms "excitation" and "inhibition" refers to the clearly defined molar concepts found in the systematic writings of Pavlov (15) and Hull (10). Eysenck postulates that personality differences between extraverts and introverts are mainly due to disturbances in the cortical excitatory-inhibitory balance (4, p. 122, 5). He suggests that overt physical and mental characteristics of anxiety, obsessions, compulsions, or ruminations are characteristic of extreme introverts (dysthymics), and these characteristics are consistent with a presumed state of strong cortical excitation and/or weak inhibition. Extreme extraverts (hysteries) are considered to be characterized by "such escape mechanisms as fugues, amnesia and gross conversion symptoms. They are typically insensitive, irresponsible, unreliable and little concerned about others." These characteristics seem to involve some form of dissociation and may reasonably be associated with a state of strong cortical inhibition and/or weak excitation. On the basis of Jung's theory and of Eysenck's cortical excitation-inhibition postulate, some possible behavioural differences between the introversive and extraversive types have been deduced, and experimentally investigated (1, 2, 5).

In this writer's opinion, the present body of knowledge about this personality dimension contains some important implications for research in the area of alcohol and alcoholism. On the basis of the scientific data and theory on this aspect of personality, the following hypotheses have been formulated relating alcohol and alcoholism to introversion-extra-version.²

¹The writer wishes to acknowledge the personal encouragement for this article from Dr. Cyril Franks, whose paper "Alcohol, Alcoholism and Conditioning," which expresses some similar views, has been published in *J. Ment. Sci.*, 1958, 104, 14-34.

²Investigation of some of these hypotheses is planned, probably to take the

ALCOHOL AND INTROVERSION-EXTRAVERSION

(1) Administration of a depressant drug (sodium amytal) is found to decrease a subject's speed of conditioning (eyeblink response) and to increase the rate of extinction (8). Alcohol is classified pharmacologically as a depressant drug. If there are no additional complex psychological factors influencing its effect on conditioning behaviour, then changes in rate of acquisition and extinction of a conditioned response similar to those noted with a depressant drug should be observed under low blood alcohol concentrations (that is, before the intoxication threshold is reached and performance is greatly disrupted or the subject loses consciousness). It may be predicted that when comparison is made between a subject's performance under moderate blood alcohol levels (.90 mg./cc. to .40 mg./cc.) and under alcohol-free conditions: (a) the rate of conditioning will be slower in the moderately alcoholized, than in the alcohol-free state; (b) the rate of extinction will be faster in the moderately alcoholized, than in the alcohol-free state.

There is evidence (5, 7) to indicate that conditioned responses in introversive subjects are more quickly acquired and more slowly extinguished than in extraversive subjects. The rate of acquisition and extinction of a conditioned response which is observed under sodium amytal appears to more closely resemble the behaviour patterns displayed by more extraversive subjects. This seems logically consistent with the presumed reduction in cortical excitation occasioned by sodium amytal, and Eysenck's postulate that extraversion is associated with less cortical excitation as compared with introversion. If this is the case, then under a depressant drug the results of the behavioural tests which Eysenck (2, 5) has claimed distinguish introverts from extraverts should be observed to shift toward more extraversive response patterns. On this inference it is hypothesized that under moderate blood alcohol concentrations: (c) a subject's physical persistence on a task will decrease; (d) his systolic blood pressure will be lower, and stressed pulse rate will decrease; (e) his responses on a test of speed and accuracy will be faster and less accurate; (f) his aspiration level for his performance will be lower and judgment of his performance will be higher; (g) his preference for simple, obviously funny, sex-concerned, and aggressive jokes will increase while preference for complex, clever jokes will decrease.

(2) Studies on the effects of a depressant drug on subjects with "known"³ degrees of introversion or extraversion suggest that introversive

following order: (1) alcoholics' relevant drinking behaviour and experiences; (2) conditioning study of alcoholics; (3) conditioning treatment of alcoholism.

³Operationally defined in terms of behavioural tests.

subjects require much larger doses of the depressant to reach the same sedation threshold (6, 16, 17). Eysenck has therefore suggested that introverted personalities, because they have greater cortical excitation and/or less inhibition, have greater tolerance for or less susceptibility to a depressant. From this work, it may be suggested that, if some arbitrary indication of alcohol intoxication (for example, word slurring) is selected, the more introverted subjects will have a higher threshold. The blood alcohol level for introverts will be greater than will the level for extraverts when this arbitrary level is reached.

(3) Since the sedation threshold for alcohol is hypothesized to be higher for introverts than for extraverts, it may also be predicted that the behavioural changes hypothesized above in (1) might be less marked for introverts than for extraverts, when similar low (.40-.90 mg./cc.) blood alcohol levels are attained.

ALCOHOLISM AND INTROVERSION-EXTRAVERSION

(1) Evidence indicates that the speed with which a conditioned response is established or extinguished is related to introversion-extraversion and is independent of neuroticism (7, 8). Since there is no basis for expecting any additional factor peculiar to alcoholics to affect the conditioned response, this relation should also be observed in an alcoholic population. Thus, it may be predicted that a conditioned response (galvanic skin response or eyeblink) will be more quickly acquired and more slowly extinguished in more introverted than in more extraverted alcoholics.

(2) If the above hypothesis is valid, then this personality dimension may offer a valuable basis on which to select alcoholics for certain treatment. The classical aversion treatment of alcoholism may be conceived of as a type of conditioning procedure.⁴ The emetic drug is the unconditioned stimulus (US), and is associated with nausea, the unconditioned response (UR). As a result of a series of acquisition trials in which alcohol, the conditioned stimulus (CS), is paired with the US to evoke the UR, the alcohol (CS) itself evokes nausea (CR).

It is assumed that, during the conditioning trials, anxiety reactions, as part of the total nausea response to the emetic drug, become conditioned to the CS of alcohol. After the CR is established and the conditioning trials have ceased, it is suggested (following the reinforce-

⁴This is intended to refer to the classical conditioning technique as described by LEMERE & VOEGTLIN (*Quart. J. Stud. Alc.*, 1950, 11, 199-204), where administration of alcohol precedes the nausea, and not to other "aversion techniques" where alcohol is administered after the patient is nauseated.

ment theory of learning) that an instrumental type of conditioning is operating to establish non-drinking behaviour. Instrumental responses which avoid the nausea (for example, not drinking, or avoiding alcohol) are strengthened by anxiety reduction. Since the anxiety is itself a learned reaction, continued avoidance of nausea results in gradual extinction of anxiety. A stronger CR of nausea could be expected to evoke stronger learned anxiety, and both should therefore show a resistance to extinction in proportion to their strength.

The strength of the conditioned nausea response may be assessed by observing the severity or duration of nausea produced by alcohol alone in the absence of the emetic drug. An indication of the strength of the learned anxiety may be obtained by observing the length of time that alcohol is avoided (that is, not consumed) after the conditioning trials cease. If introversive alcoholics are characterized by faster acquisition and slower extinction of a conditioned response and the alcohol→nausea response is established, this response should be more strongly established and should show greater resistance to extinction among the introversive alcoholics than among the extraverted ones. It is therefore predicted that after patients who have had similar conditioned aversion treatment for alcoholism cease this treatment, the more introversive patients will have a longer average length of abstinence than will the more extraversive patients.

While the above hypothesis is based only on the findings from conditioning studies on introverts and extraverts, Jung's theory of introversion-extraversion itself might also suggest this prediction. In contrast to the extravert, the introvert is assumed to be greatly concerned about maintaining a subjective security in his environment. He is more thoughtful, cautious, and concerned about anticipating and predicting the consequences of his actions. It seems likely, therefore, that the relation between drinking alcohol and suffering physiological illness would be more seriously considered by the introvert. With weak cortical inhibition, or strong cortical excitation, this alcohol→nausea response by the introvert might be elicited more readily, and for a longer period of time. As a result, this relation should deter drinking more strongly, or for a longer period of time for the introvert than for the extravert.

(3) From the personality characteristics assumed to be related to extraversion, an hypothesis about the kind of treatment most suited to extraversive alcoholics may be formulated. Extraverts are assumed to be more inclined to act out than to internalize their feelings, and to be typically less reflective and more impulsive. They quickly adjust to new surroundings and move among objects and people with relatively little caution or fear. They are much less keenly aware of themselves and their

individuality. From this description it would appear that extraverts might prefer, and do better under, treatment which was on a less intellectual and individual basis, and involved more emotional acting out of feelings. "Alcoholics Anonymous" emphasizes the strong "common-tie" group feeling by combining the members' own recounted personal drinking histories with their emotionally experienced conversion and cure. This type of activity would probably be more suited to extravertive than to introvertive individuals. The latter would probably feel more comfortable in a clinic setting where treatment was designed on a more individual and intellectual basis. Thus, it is predicted that in the group of alcoholics who have experienced both A.A. and clinic treatment for alcoholism, the more extraverted patients will feel more at ease with A.A., and feel that it has been of greater help than the clinic in curing their drinking problem. The contrary should be true of the more introverted alcoholics.

(4) In Jellinek's phases of the drinking history of alcoholics (11), the Prodromal Phase begins when a "blackout" occurs after ingestion of a relatively small amount of alcohol. This phase, it is said, may last for a period of months or years, and is terminated, by Jellinek's definition, when these blackouts become "frequent."⁵ A blackout is defined as amnesia of a few hours duration, for one's activities or experiences. Blackouts thus defined may be thought of as a type of mental dissociation which might then also be theoretically linked with extraversion and cortical inhibition or under-excitation. Since alcohol is a depressant drug, and the blackout phenomena result upon alcohol ingestion, the occurrence of these blackouts, as defined by Jellinek, may be related to the alcoholic's position on the scale of introversion-extraversion. On the assumption that introvertive personalities have greater tolerance for larger quantities of alcohol, it is predicted that: (a) a smaller proportion of the drinking histories of introvertive alcoholics will report the occurrence of any blackout phenomena, as compared with the histories of extravertive alcoholics; (b) the Prodromal Stage, whose onset is marked by the first blackout and whose termination is indicated by an "increasing frequency" of these blackouts, will be significantly longer for introvertive than for extravertive alcoholics.

In Jellinek's investigation of drinking phases, the occurrence of solitary drinking was reported by the majority of the alcoholics, but this behaviour did not appear consistently at any stage in the drinking histories. It has been suggested, therefore, that the onset of solitary drinking is more affected by personality and situational factors than by

⁵"Increasing frequency" is defined by Jellinek (12) to be "at least two or three times out of ten drunks."

what is common to the natural development of the alcoholic drinking pattern. Usually this behaviour is considered to indicate a movement toward social isolation, occasioned by the alcoholic's increasingly poor psychological adjustment or difficulties in inter-personal relations. It seems equally possible, however, that solitary drinking may be indicative of an introversive personality. Eysenck states that, in contrast to the extravert, "... the introvert does not particularly care to be with people, would rather be alone..." (3, p. 121). If the need arises, both introverts and extraverts can effectively take part in social situations, for the degree to which inter-personal relations are adequately formed is related rather to the neurotic aspect of personality. Studies (3, 9) of the social behaviour of people with known degrees⁶ of neuroticism and extraversion give some support for Eysenck's claim that the introversive preference for solitude is independent of neuroticism. It may be expected, therefore, that among alcoholics the introverts, typically preferring solitude, would more frequently report solitary drinking as compared with the extraverts. Solitary drinking in the latter group might occur when drinking had become a sufficient problem to disrupt social relations and might therefore be more indicative of neurotic social shyness. If provision is made to control the neuroticism factor, it may be predicted that: (c) a greater proportion of the drinking histories of introversive alcoholics will report solitary drinking behaviour as compared with the histories of extraversive alcoholics.

It also seems possible that introversion-extraversion may be related to steady and periodic drinking in alcoholism. It has been suggested that the introvert would be more prone to insomnia, and alcohol might be used by him as a night-cap to "put his mind at rest" before retiring. Alcohol may also be employed differentially to reduce tension in social situations since the introvert is at times somewhat ill at ease with people, and acutely aware of himself. The introvert, however, is also typically introspective, cautious, and conscientious, and he might as a result be less likely to drink great amounts at one time, or to have irresponsible, impulsive drinking "sprees" or binges. In contrast, the extravert, characteristically being relatively irresponsible, insensitive, and unconcerned about others, may be expected to drink in considerable quantities, and in impulsive spreeds. Just as a normal introvert might be expected to drink frequently, but in small amounts, an alcoholic introvert might be a "steady drinker" (that is, one who drinks more or less the same amount at regular, frequent intervals). A normal extravert, by contrast, can be expected to drink larger amounts spasmodically, and, by analogy, the

⁶Neuroticism and introversion are defined in terms of test scores on previously validated questionnaires (3).

alcoholic extravert might be expected to be the "periodic drinker" (that is, a person who does most of his drinking in bouts of two or three or more days, either not drinking at all between bouts, or drinking only very moderately). On this basis it is hypothesized that: (d) introversive alcoholics will be found to be mostly steady drinkers while extraversive alcoholics will be mostly periodic drinkers.

(5) Leucotomized patients or individuals with known organic brain damage are found to present markedly extraverted behaviour patterns (that is, slow acquisition, fast extinction of conditioned eyeblink). Studies on alcoholics frequently suggest a causal link between long-term chronic alcoholism and progressive organic brain damage. If this is the case, a greater number of extraverted test scores may be obtained in a group of long-term chronic alcoholics, as compared with a group of alcoholics having only a few years of alcoholism.

SUMMARY

On the basis of the personality dimension of introversion-extraversion, as conceptualized and operationally defined by Eysenck, hypotheses have been derived which relate certain aspects of alcohol and alcoholism to introversion-extraversion. These hypotheses are presented with two aims in mind: first, that of extending and evaluating an application of Eysenck's personality theory to the area of alcohol and alcoholism, and secondly that of suggesting a fertile, testable, theoretical framework within which a systematic examination of alcohol and alcoholism may be made.

REFERENCES

1. EYSENCK, H. J. *Dimensions of personality*. London: Routledge & Kegan Paul, 1947.
2. EYSENCK, H. J. *The structure of human personality*. London: Methuen, 1953.
3. EYSENCK, H. J. The questionnaire measurement of extraversion and neuroticism. *Rivista di Psicologia*, 1957, 113-140.
4. EYSENCK, H. J. Drugs and personality: Theory and methodology. *J. ment. Sci.*, 1957, 103, 119-131.
5. EYSENCK, H. J. *Dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.
6. EYSENCK, H. J., CASEY, S., & TROUTON, D. S. Drugs and personality; effect of stimulant and depressant drugs on continuous work. *J. ment. Sci.*, 1957, 103, 645-649.
7. FRANKS, C. M. Conditioning and personality; a study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, 52, 143-150.
8. FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, 48, 119-125.
9. FRASER, R. *The incidence of neurosis among factory workers*. London: H.M. Stationery Office, 1947.

10. HULL, C. L. *The principles of behavior*. New York: Appleton-Century-Crofts, 1943.
11. JELLINEK, E. M. Phases in the drinking history of alcoholics. *Quart. J. Stud. Alc.*, 1946, 7, 1-88.
12. JUNG, C. G. *Psychological types*. London: Kegan Paul, Trench, Trubner, 1923.
13. JUNG, C. G. *Contributions to analytic psychology*. New York: Kegan Paul, 1928.
14. McDOUGALL, W. The chemical theory of temperament applied to introversion and extraversion. *J. abnorm. soc. Psychol.*, 1929, 24, 293-309.
15. PAVLOV, I. P. *Conditioned reflexes and psychiatry*. London: Lawrence and Wishart, 1941.
16. SHAGASS, C. The sedation threshold. *EEG Clin. Neurophysiol.*, 1954, 6, 221-225.
17. SHAGASS, C., & NAIMAN, J. The sedation threshold as an objective index of manifest anxiety in psychoneurosis. *J. Psychosom. Res.*, 1956, 1, 49-57.

A NON-PARAMETRIC APPROACH TO THE GRAPHICAL ANALYSIS OF TRENDS

RICHARD H. WALTERS

University of Toronto

PSYCHOLOGISTS FREQUENTLY WISH to examine the responses of a group of subjects over a series of trials on a task. The usual method of handling the data is to sum the scores of the subjects for each trial and then find the trial means. Results are then graphically represented by drawing a curve which best fits these trial means. In addition, a test of the significance of differences between trial means is sometimes carried out by means of Fisher's analysis of variance technique. However, the data obtained from psychological experiments do not always meet the requirements of a parametric analysis. In this event, a non-parametric analysis can, and should, be used.

Let us suppose that R subjects have been given k trials on a particular task. Thus, there are R sets of k measures. The measures in each of these R sets are now ranked from 1 to k , the smallest measure being given the rank of 1. These ranks may then be summed over the rows, that is, the subjects, to give a sum of ranks in each column, that is, for each trial. A graphical representation of results may be produced by plotting sums of ranks against trial numbers. The Friedman two-way analysis of variance by ranks, as outlined by Siegel (2), can be used to determine whether results differ significantly from trial to trial.

The procedure can be illustrated by reference to an experiment by Jakubczak and Walters (1). Twenty-four children were exposed to the autokinetic effect over a series of eight trials and on each trial were asked to judge how far the light moved. Whatever the size of their judgment, the experimenter's confederate made a judgment 5 in. greater than that made by the subject. Under these conditions, not only were the judgments of individual children extremely diverse on any single trial, but some children made an occasional erratic judgment, for example, 100 in., which could be no more than a wild guess induced by the continued suggestion. Consequently, a non-parametric analysis of the trend of the subjects' responses from trial to trial was undertaken.

First, the eight responses made by each subject were ranked among themselves. Subject no. 2 gave the following series of responses: 0, 15, 12, 17, 12, 14, 24, 18. These responses were transformed into ranks:

1, 5, 2.5, 6, 2.5, 4, 8, 7. The same procedure was followed with the responses of all 24 subjects. The sums of the ranks of the 24 subjects were then calculated for each of the eight trials. For Trials 1 through 8 the sums were as follows: 52.5, 80.0, 74.0, 87.0, 110.0, 116.0, 117.5, 118.0. A graph illustrating the trend of responses could then be set up with the trial numbers located at appropriate places on the abscissa and the sums of ranks on the ordinate.

In addition to avoiding many of the assumptions involved in the use of parametric techniques, this procedure has the advantage that the responses of all subjects are given precisely the same weight over the total series of trials, that is, the sum of the ranks of each individual subject is equal to $k(k+1)/2$, where k equals the number of trials. The sum of the column (trial) totals should, of course, equal $R \times k(k+1)/2$.

In experiments such as the one referred to above, a response may not be available for all subjects on all trials. If the gaps are not too numerous, these can be filled by inserting the subject's median response within his series of trials before ranking. The trial to which the median response is assigned should probably be determined by random selection. In this case, the position of other responses in the series will require adjusting accordingly. For example, if the median has been assigned to Trial 2, the subject's response to this trial will now be regarded as his response to Trial 3, and so on. Substitution of this kind should, of course, be avoided if possible; there are times, however, when some loss of data cannot be prevented.

REFERENCES

1. JAKUBCZAK, L. F. & WALTERS, R. H. Suggestibility as dependency behavior. *J. abnorm. soc. Psychol.* (In press.)
2. SIEGEL, S. *Non-parametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

A NOTE ON LAMBERT'S "EVALUATIONAL REACTIONS TO SPOKEN LANGUAGES"

HENRI TAJFEL

*University of Oxford*¹

LAMBERT *et al.* (1) have recently reported a study which is of uncommon interest for our understanding of the functioning of stereotypes. The purpose of this note is to point out some aspects of the study which seem to be of special theoretical importance, and to show that the results, though they appear to be unexpected in some ways, are consistent with general considerations about the nature of shifts in judgments.

Lambert's subjects were groups of French-speaking and English-speaking Montreal students. They were asked to evaluate the personality characteristics of four bilingual speakers who recorded on tape French and English versions of a 2½-minute passage of prose. The subjects were not aware of the fact that each speaker read the passage in both languages, "... so that the evaluational reactions to the two language guises could be matched for each speaker." The cumulative results for all the speakers showed that the English subjects evaluated the English speakers more favourably on seven out of fourteen traits; the French subjects evaluated the *English* speakers more favourably on ten out of fourteen traits. Already at this general level, and without further analysis, this finding is of interest: it contradicts the oversimplified view that national stereotypes are determined by an autistic, uncritical, and wish-fulfilling image of one's own group, especially when this group is contrasted with another in a context of latent or explicit tension or conflict.

Lambert *et al.* examine and reject a number of explanations which might account for the French preference of the English speakers:

(1) The possibility that the lower rating of the French speakers by the French subjects was due to the selection by the experimenters of traits which did not have "value" for French-speaking Canadians. This does not meet the facts, as some of the traits used were rated as highly desirable by the French group of subjects.

(2) The possibility that the results were due to "... the greater

¹Now in the Department of Social Relations, Harvard University, on a year's leave of absence from Oxford.

probability in the Montreal community of finding English people in more powerful social and economic positions," and therefore to the existence of powerful stereotypes common to both sections of the community. In addition to cogent arguments presented by Lambert *et al.* to reject this alternative, it should be noted that it would not explain the emergence in the data of the fact that the French group makes a greater use of these stereotypes than the English group.

(3) The possibility that data from various questionnaires designed to elicit the subjects' attitudes towards their own and the contrast groups might yield some convincing correlations with the ratings of voices. Lambert *et al.* conclude "... that the comparatively unfavourable perception of French speakers is essentially independent of the perceivers' attitudes towards French and English groups."

The authors tentatively accept the possibility that we are dealing here with a phenomenon akin to what has sometimes been called "self-hatred" among the Jews: the adoption by a minority group of stereotyped attributes assigned to it by the majority. This explanation runs into a number of difficulties. First, "self-hatred" has been shown to exist in situations involving immeasurably more tension and conflict than the intergroup situation in Montreal: in Nazi Germany, concentration camps, etc. Secondly, it has not yet been shown that the adoption of the majority norms may also involve an exaggeration of their value. The "self-hating" Jew may take over from the Nazis their evaluation of his ethnic group; but he does not make it even more unfavourable. According to Lambert's findings, the French Canadians do not just accept the majority stereotype of them: they do it with a vengeance. The authors quote French-Canadian descriptions of the French and English groups, given in an open-ended questionnaire, as some evidence for the "self-hatred" hypothesis. These descriptions reflect ambivalent attitudes of the French Canadians towards both ethnic groups. This, however, cannot be considered to support directly the application of the hypothesis to the ratings of voices. There is no way of explaining on this basis why the "self-hatred" should cause consistent relative underrating by the French of their own group on some particular traits, and not on others—and this, as will be seen below, is the main drift of the findings. One might also perhaps add a guess that ambivalent attitudes as mild as those described by Lambert *et al.* could be elicited from any national group not hopelessly lost in a fit of megalomania.

Some hypotheses concerning shifts of judgments in social situations have been recently advanced by the writer (5). Two assumptions were at the basis of predictions which applied to a number of situations where judgments are made along a continuum, physical or abstract.

The first assumption was that in a series in which the value of the stimuli to the subject was correlated with the physical dimension under judgment (for example, a series of coins judged in terms of size), the differences between the stimuli of the series would be judged larger than in a physically identical series not displaying a correlation between physical magnitude and value. A few recent studies, and some earlier ones, provide experimental evidence for this assumption (4, 6, 7). Secondly, it was postulated that when a discontinuous classification, correlated with the dimension being judged, is superimposed on a series of stimuli, judgments of stimuli falling into the distinct classes would be shifted in the directions determined by the class membership of these stimuli. This can be visualized as a rubber band stretched outwards from its middle in both directions. Industrial products originating from two sources and differing consistently, according to their source, in size, weight, colour, texture, etc., provide one set of examples for such series. Shifts in the directions consistent with the classification are predicted to occur when the stimuli are labelled as originating from one source or another *before* the judgment of physical magnitude is made.

A third assumption was derived from the first two: namely, that when the classification superimposed on the series is of inherent value or relevance to the subject, the shifts of judgment should be in the same direction as in the series just discussed, but more pronounced. Some evidence for this has been provided in the recent studies by Secord, Bevan, and Katz (3), and by Pettigrew, Allport, and Barnett (2).

From this point of view our main interest is in the judged *differences* in various traits between the French and English guises of the speakers in the Lambert *et al.* study. These differences provide a clue to the finding that on some traits the French subjects, as compared with the English subjects, underrate the French, or overrate the English, speakers. There exists in the Montreal community a discrepancy in socio-economic status in favour of the English group. Both groups of subjects are aware of this: when estimating the likely occupations of the speakers, they ascribe significantly higher status to the English than to the French guises.

This in itself does not account for the findings for reasons already discussed. However, the fact is established that the classification into French and English is correlated with socio-economic status, both objectively and subjectively. The subjects judge the speakers on a number of traits: some of these traits may be related to socio-economic status or success, some not. So far, the situation is identical for the French and English groups of subjects. The only difference between the groups is

in the relevance to them of the Franco-English discrepancy. It is a fair assumption that this discrepancy causes more concern, is in a sense more salient, worrying, and relevant to the French than to the English subjects, especially as the French subjects were all college students, future direct competitors of the English group. If this is so, *some* differences between the French and the English are of greater impact to them than to their English counterparts. Therefore, the prediction could be made that the classification into French and English would determine larger shifts in both directions for the French group on those dimensions which are correlated with the "value" or relevant aspect of this classification—the socio-economic status.

The comparison between the judgments of the French and the English subjects should show that (a) the French subjects tend to accentuate more than the English subjects the English superiority on traits related to socio-economic success; (b) the French subjects, as compared with the English subjects, should not show this trend on traits not related to socio-economic success.

The comparison *within* the French group of judgments on traits related to socio-economic success with those on traits not related to it, should show greater accentuation of differences in favour of the English on the former than on the latter.

Lambert *et al.* report that the English of their four speakers was "faultless." As to their French, two of them (Cou and Bla) spoke with a French-Canadian accent, one (Leo) "spoke with a marked French-Canadian accent characteristic of those who work in the 'bush'" (they describe it later as a "caricatured French Canadian"), and the fourth (Tri) "spoke French with an accent that was judged indistinguishable from that used in France." If the French of the French guises was identified in this way by the French group of subjects, the inference from the predictions just stated would be that the French subjects would accentuate the differences in the relevant traits between the French and English guises in their judgments of the first two speakers; they should still be concerned, but perhaps less so, with the "bush" accent; and not at all concerned with the "Parisian" accent.

Table I in the Lambert *et al.* study provides all the data needed to assess these inferences. It contains quantitative statements of significance of differences in evaluations of each trait for each speaker between his French and English guises for both groups of subjects.

These data have been reclassified (see Table I) into eight classes. The first six columns contain traits for which there are significant differences between the judgments of the English and French guises of the same speaker, the last two the non-significant ones.

TABLE I
A RECLASSIFICATION OF DATA FROM THE LAMBERT ET AL. STUDY (SEE TEXT)

Speaker	A	B	C	D	E	F	G	H
COU	Height Good looks Character		Kindness Likeability	Leadership Sense of humour Intelligence Self-confidence Dependability Ambition Sociability			Leadership Sense of humour Intelligence Religiousness Self-confidence Dependability Entertainingness Ambition Sociability	Religiousness Entertainingness Kindness Likeability
BLA	Height Good looks Ambition		Kindness	Leadership Intelligence Self-confidence Dependability Sociability Character		Religiousness Kindness	Leadership Sense of humour Intelligence Religiousness Self-confidence Dependability Entertainingness Sociability Character Likeability	Sense of humour Entertainingness Likeability
LEO	Height Good looks Intelligence Dependability Ambition Character		Sense of humour	Leadership Self-confidence Sociability Likeability		Kindness	Religiousness	Sense of humour Religiousness Entertainingness
TRI		Height Intelligence Dependability			Sense of humour Entertainingness Sociability	Self-confidence	Self-confidence Entertainingness Kindness Sociability Likeability Leadership	Height Good looks Leadership Sense of humour Intelligence Religiousness Dependability Entertainingness Kindness Ambition Character Likeability

The columns contain the traits on which:

- A. Both groups of subjects judged the English guise of a speaker more favourably than his French guise.
- B. Both groups of subjects judged the French guise more favourably.
- C. English group judged the English guise more favourably.
- D. French group judged the English guise more favourably.
- E. English group judged the French guise more favourably.
- F. French group judged the French guise more favourably.
- G. In the English group no significant differences were noted between the English and French guises of a speaker.
- H. In the French group no significant differences were noted between the English and the French guises of a speaker.

As can be seen from Table I, the patterns for Bla and Cou, the two French-Canadian accents, are almost identical. Both groups agree that the English guises are better looking and taller. However, the French group has for both a cluster of traits (leadership, intelligence, self-confidence, dependability, sociability) clearly related to socio-economic success in which it judges the differences between the French and the English guises to be significantly in favour of the English; for the English subjects all these traits can be found in the non-significant column (G). To this can be added "character" for Bla where the difference is significant for the French subjects and not significant for the English; and, in the same way, "ambition" for Cou. For Bla, the French guise is at the favourable end for the French subjects on the traits of religiousness and kindness, not related in any clear manner to socio-economic success. For Cou, these traits are in the non-significant column (H).

The pattern for Leo, the "bush" accent, is similar, apart from two differences: the French subjects significantly dislike him—or like him less than his English counterpart; and the English subjects join the French in estimating his intelligence and dependability to be significantly lower than that of his English counterpart. This might be due to the fact that some of the English subjects spoke at least some French, and were able to recognize Leo's "bush" French accent for what it was.

The case of the "Parisian" Tri is strikingly different. The French subjects do not think that the English are better at anything than he is. The entire "success" cluster travels from column D to column H. A comparison between the judgments of the French and the English subjects shows that the differences in these traits are not relatively accentuated in favour of the English guise by the French subjects.

In summary, it seems that the hypothesis of accentuated differences in the judgment dimensions relevant to a value classification does account

for the results of the Lambert *et al.* study. For the French group of subjects the differences between the English and the French in those traits which are relevant to socio-economic status are more pronounced than the same differences for the English group of subjects; these differences are in the direction, consistent with the classification, of *relatively* overrating the English or underrating the French. No such tendency is shown by the French subjects (a) for traits not relevant to socio-economic status (kindness, likeability, religiosity, entertainingness); (b) for traits relevant to socio-economic status, but inherent in an individual who is not "one of them" in the Franco-English competition: a Frenchman from France.

REFERENCES

1. LAMBERT, W. E., HODGSON, R. C., GARDNER, R. C., & FILLERBAUM, S. Evaluational reactions to spoken languages. *J. abnorm. soc. Psychol.* (in press).
2. PETTIGREW, T. F., ALLPORT, G. W., & BARNETT, E. O. Binocular resolution and perception of race in South Africa. *Brit. J. Psychol.*, 1958, 49, 265-278.
3. SECORD, P. F., BEVAN, W., & KATZ, B. The Negro stereotype and perceptual accentuation. *J. abnorm. soc. Psychol.*, 1956, 53, 78-83.
4. TAJFEL, H. Value and the perceptual judgment of magnitude. *Psychol. Rev.*, 1957, 64, 192-204.
5. TAJFEL, H. Quantitative judgment in social perception. *Brit. J. Psychol.*, 1959, 50, 16-29.
6. TAJFEL, H. The anchoring effects of value in a scale of judgments. *Brit. J. Psychol.* (in press).
7. TAJFEL, H., & CAWASJEE, S. D. Value and the accentuation of judged differences: A confirmation. *J. abnorm. soc. Psychol.* (in press).

TRAITS AND COLLEGE ACHIEVEMENT

D. D. SMITH

McGill University

PREVIOUS PAPERS (6, 7) have described the factorial analysis of a group of ability and interest measures, the definition of the factors as traits of personality, and the validation of these traits by reference to data concerning choice of undergraduate study programme and achievement in the first year of college. These papers are referred to as the "1951 study" in this paper.

The aims of this study were two: (a) to confirm, if possible, the existence of the composite interest-ability factors, defined as traits in (6), through the analysis of a somewhat different group of measures; (b) to employ the relations between trait scores and choice of study programme and achievement, demonstrated in (7), in the development of criteria for use in the selection of applicants for conditional undergraduate standing in the Evening Division of Sir George Williams College, Montreal.

PROCEDURE

The measures included were the verbal reasoning, numerical ability, and abstract reasoning tests, Form A, of the *Differential Aptitude Tests* (1), the *Kuder Preference Record-Vocational, Form BB* (4), the *Nelson-Denny Reading Test, Form A* (5), the *Survey of Study Habits and Attitudes* (2), and the *Gordon Personal Profile* (3). These are referred to in Table II as DAT, KV, ND, SSHA, and GPP, respectively.

The first four measures were retained from the 1951 study. It is generally accepted that verbal fluency and comprehension represent one of the most important single variables in academic achievement. The Nelson-Denny Reading Test, yielding a vocabulary score and a reading comprehension score, was included for this reason. Although personality adjustment has proved to be a rather elusive area to measure through paper and pencil group methods, experience in counselling leads to frequent assertions of its relevance to academic achievement. The Gordon Personal Profile, a forced-choice paper and pencil inventory yielding four scores purporting to measure ascendancy, responsibility, emotional stability, and sociability, seemed to show more promise than many instruments for assessing this area. Finally, if it is true that emotional maladjustments may, in part, express themselves through attitudinal postures, then a measure of study attitudes should show relations both with achievement and with measures of personal adjustment. The Survey of Study Habits and Attitudes, another inventory, yields a single score which is shown in the manual to give a weighted average correlation of .41 with one-semester grade averages, for samples totalling 1,249 students drawn from five United States universities, and a

weighted average correlation of .25 with the American Council on Education Psychological Examination for the same groups. These measures yielded a total of 19 scores (total scores on the Nelson-Denny and Gordon Personal Profile were not included).

This battery was administered to 255 Day College freshmen at Sir George Williams in the fall of 1956. Product-moment correlations were computed for every pair of variables and this matrix of coefficients was analysed by the complete centroid factor analysis method (8). The original correlations ranged in magnitude from .66 to -.51. Eight factors were extracted, after which the residual coefficients were normally distributed, ranging in value from .13 to -.21, with a mean value of -.013 and a modal value of .01.

TABLE I
FINAL ROTATED OBLIQUE FACTOR MATRIX $V_{R\uparrow}$

Variable	Factor							
	A	B	C	D	E	F	G	H
DAT Numerical	12	38*	51*	08	10	01	02	-16
DAT Verbal	63*	12	20*	-04	-04	-17	-02	04
DAT Abstract	36*	43*	37*	-15	14	-25*	-12	01
KV Mechanical	22*	65*	-02	06	03	07	-59*	02
KV Computational	-05	02	11	09	57*	20*	28*	-20*
KV Scientific	10	53*	06	06	02	51*	11	06
KV Persuasive	14	-04	27*	-13	23*	-32*	-38*	05
KV Artistic	06	03	-10	-02	06	-01	-07	56*
KV Literary	-01	-58*	-21*	23	-25*	-04	39*	-25
KV Musical	-20*	-29*	-09	-05	-19	-46*	06	-19
KV Social Service	-16	-28*	-05	09	-48*	10	-16	-07
KV Clerical	-03	-11	-19	06	66*	00	10	-17
GPP Ascendancy	15	-24*	80*	04	-07	26*	21*	40*
GPP Responsibility	03	-14	04	72*	-01	22*	06	10
GPP Emot. Stability	02	-15	04	62*	02	01	-01	15
GPP Sociability	-04	-10	65*	01	-21*	13	-04	13
SSHA	09	-21*	05	50*	-08	39*	15	06
ND Vocabulary	63*	-16	-02	09	-09	00	06	03
ND Paragraph	74*	-04	06	00	00	01	-03	13

†Decimal points omitted. *Significant loading.

The factors were rotated by the method of two-dimensional sections until simple structure emerged. The cosines between the final rotated factors ranged in value from .40 to -.49, with 14 of the 21 cosines having an absolute value equal to or smaller than .20. The factorial structure is, therefore, oblique.

IDENTIFICATION OF THE FACTORS

These factors, identified in terms of the variables with significant loadings (see Table I), may be named and described as follows:¹

¹A significant loading was arbitrarily defined as a loading of absolute value equal to or greater than .20, equivalent to the $P = .001$ level of significance for zero order correlation coefficients where $N = 255$.

Factor A: Verbal ability. This factor, extending in a positive plane only, can be assumed to be a measure of verbal ability much like Factor B of the 1951 study.

Factor B: Scientific creativity vs. Aesthetic creativity. This, a bipolar factor, contrasts preferences for scientific professions and inventive and manipulative tasks involving physical objects and objective facts with preferences for occupations in the creative arts and inventive and manipulative tasks involving music, literature, etc. The numerical ability and abstract reasoning tests both show substantial positive loadings.

Factor C: Self-confidence. A second factor with positive extension only, this involves responses indicating active, self-assured, assertive relations in groups; a tendency to make independent decisions; enjoyment of dominant inter-personal roles; and ease in influencing others. The numerical, verbal, and abstract tests all show significant positive loadings on this factor, suggestive of the role that above average intellectual endowment may play in the development of this aspect of personal adjustment.

Factor D: Emotional maturity. This factor, with positive extension, involves responses indicating a stable, persevering, determined approach to responsibility; freedom from undue tension and anxiety; and relative freedom from poor study habits and unfavourable study attitudes.

Factor E: Accounting vs. social service. This, a bipolar factor, opposes preferences for orderly-systematic accounting and clerical activities at the positive pole with preferences for social service type occupations and activities at the negative. In its positive pole it resembles Factor D of the 1951 study.

Factor F: Objective observer vs. subjective performer. Another bipolar factor, this contrasts preferences for objective, scientific professions and tasks (positive pole) with preferences for occupations and tasks, such as those of the radio singer, sales manager, salesman, public speaker, etc., which appear to put the emphasis on the performer and subjective interpretation. This factor may be considered identical with Factor C of the 1951 study. Although it bears a resemblance to Factor B, described above, there are several distinctions between them. First, the cosine between the reference vectors defining these factors took a value of $-.34$ indicating that while there was some convergence it was by no means as much as the verbal factor descriptions would suggest. Second, while both the numerical and abstract tests showed significant positive loadings on Factor B, only the Abstract test is related to Factor F, and the loading in this case is negative. Third, neither the mechanical nor the literary scores on the Kuder Preference Record, which show high loadings on Factor B, are of any significance in Factor F.

Factor G: Manipulate symbols vs. manipulate objects. A bipolar factor, this is defined by preferences for occupations and tasks involving the manipulation of words and numbers (Kuder computational and literary scores) as opposed to preferences for the manipulation of objects and people (Kuder mechanical and persuasive scores). Its counterpart in the 1951 study appears as Factor E.

Factor H: Create with materials vs. create with symbols. Responses indicating self-assurance and assertion, and preferences for occupations such as those of the artist, architect, sculptor, and portrait painter define the positive pole of this factor. The negative pole is defined by responses showing passivity and a lack of self-confidence, and preferences for symbol creative tasks such as those of the poet, writer, professor of mathematics, and other men working mainly with ideas.

TABLE II
MEAN TRAIT SCORES AND STANDARD DEVIATIONS FOR THE 1956 SAMPLE, WITH CRITICAL RATIOS FOR SUB-SAMPLE MEAN SCORE DIFFERENCES

Trait	Total		Arts		Science		Commerce		Critical ratios for sub-group differences		
	\bar{X}	S.D. [†]	\bar{X}	S.D.	\bar{X}	S.D.	\bar{X}	S.D.	Arts-Sci.	Arts-Com.	Sci-Com.
A	0.25	7.71	2.91	6.58	-1.80	8.14	-0.48	7.53	4.28*	3.05	1.08
B	0.19	7.65	-5.49	6.08	5.69	6.11	0.47	5.90	12.42*	6.34*	5.55*
C	0.06	8.24	-0.09	8.67	-0.41	7.61	0.81	8.37	0.26	0.67	0.17
D	0.09	7.63	-1.06	7.25	1.03	8.28	0.32	7.05	1.80	1.24	0.60
E	0.12	7.75	-4.24	7.34	-0.03	5.63	5.55	7.00	4.34*	8.82*	5.58*
F	0.13	6.73	-2.33	5.67	3.60	6.78	-1.08	6.10	6.30*	1.34	4.68*
G	0.06	7.50	3.00	7.26	-2.88	6.43	0.07	7.59	5.76*	2.52	2.66
H	0.06	7.20	0.47	7.61	2.37	6.45	-2.08	6.75	2.70	1.45	4.32*

[†]These standard deviations should not be confused with the arbitrary standard deviations of 10.0 which were used to solve the trait equations. The above deviations are those of the estimated trait scores.

*Significant at the .001 level.

APPLICATION

Multiple regression equations were developed, using the factor loadings of the measures with significant loadings as the entries for the dependent variable, to provide equations for calculating trait scores from the raw scores on the various measures. For purposes of solving these equations the independent variables (trait scores) were arbitrarily assigned means of 0.0 and standard deviations of 10.0. These equations were then applied to the raw scores for each individual in the sample so that for each a set of eight trait scores was obtained. Information was also made available on the faculty in which each student was registered and on his achievement at the end of the first year of study.

Product-moment correlations relating scores on each trait to scores on every other trait, and to total grade point and grade point average were determined for the entire sample and for the Arts ($N = 90$), Science ($N = 90$), and Commerce ($N = 75$) sub-samples. Means and standard deviations for each trait and for total grade point and grade point average were also calculated for these groups. The results of these calculations are shown in Tables II and III.

Examination of Table II shows that differences in mean scores on six of the eight traits are significant at the $P = .001$ level for at least one pair of faculty sub-groups in each case. The two traits which show no significant inter-faculty differences are C (self-confidence) and D (emotional maturity).

Table III shows the coefficients between the trait scores and both total grade point and grade point average. Although highly interrelated, these two criteria of academic achievement were used because it was felt that grade point average did not reflect the course load that a student might be carrying. Total grade point retains this feature, an important consideration in the admission of conditional undergraduates. It will be observed that five of the traits showed correlations with at least one of the criteria for at least one of the sub-groups significant at the $P = .01$ level or better. Generally, there were a smaller number of significant correlations with total grade point than with grade point average, but the former tend to be of larger absolute magnitude than the latter. Traits A (verbal ability), D (emotional maturity), F (objective observer *vs.* subjective performer), and H (create with materials *vs.* create with symbols) show the highest numbers of significant relations. As Table III shows, the interrelations of these four traits are very small.

Multiple regression coefficients were calculated for each of the three sub-groups for estimating total grade point. In each case the four traits showing the highest absolute correlation with the criterion were used as

TABLE III
PRODUCT-MOMENT CORRELATION COEFFICIENTS FOR TRAIT INTERRELATION, AND
RELATIONS OF TRAITS TO TOTAL GRADE POINT AND GRADE POINT AVERAGE FOR 1956
SAMPLE†

Trait		A	B	C	D	E	F	G	H	Total Grade point	Grade Point average
A	Total		-08	20	07	28	00	00	00	24**	21**
	Arts		13	22	18	29	06	-16	09	30**	20
	Science		07	25	20	21	11	00	02	20	12
	Commerce		11	16	-12	21	11	-14	-04	33**	27*
B	Total			09	08	17	47	-57	24	07	09
	Arts			20	16	02	23	-49	16	04	06
	Science			20	14	26	39	-52	28	10	03
	Commerce			-07	-35	14	41	-47	15	-13	14
C	Total				21	01	-02	-14	42	02	03
	Arts				21	12	-07	-32	48	11	04
	Science				16	01	00	-17	49	-03	09
	Commerce				30	06	05	06	35	02	04
D	Total					09	43	10	10	23**	19**
	Arts					04	43	07	03	20	18
	Science					02	54	14	15	25*	21*
	Commerce					18	23	25	09	23*	24*
E	Total						02	17	-19	-11	06
	Arts						01	32	-31	-19	11
	Science						10	03	-02	-08	06
	Commerce						-16	52	-11	23*	24*
F	Total							00	00	25**	12*
	Arts							24	-10	27**	13
	Science							13	18	30**	27**
	Commerce							04	02	00	03
G	Total								-47	03	16**
	Arts								-53	02	07
	Science								-44	07	18
	Commerce								-39	16	19
H	Total									-06	21**
	Arts									-03	22*
	Science									-14	21*
	Commerce									-28*	27*
Total Grade Point											
	Total										84**
	Arts										72**
	Science										93**
	Commerce										93**

†Decimal points omitted. *Significant at the .05 level. **Significant at the .01 level.

the independent variables. The coefficients obtained were .493 for Arts, .408 for Science, and .523 for Commerce. These coefficients are all significant at better than the $P = .001$ level. Simple additive formulae were also constructed for estimating most suitable faculty of registration. These formulae and the results of their application to the total sample and each of the sub-groups are shown in Table IV.

TABLE IV
APPLICATION OF 1956 FACULTY REGISTRATION FORMULAE TO 1956
DAY DIVISION FRESHMEN

Sample	N	Formula for		
		Arts (A - B - E + G) \bar{X}	Commerce (2E - 2H) \bar{X}	Science (B + F - G + H) \bar{X}
Arts	90	15.65	-11.29	-7.56
Science	90	-10.34	14.54	-4.82
Commerce	75	-6.39	-2.79	15.26
Total	255	0.00	0.32	0.12

DISCUSSION

It is to be noted that four of the factors of this study (A, E, F, and G) have their counterparts in the 1951 study although there is a substantial difference in the batteries of measures being analysed in the two studies. (Twelve measures are common to the twenty-four of the first analysis and the nineteen of the second.) It may also be observed that four of the eight factors emerging from this study (A, B, C, and F) are defined in terms of both ability and preference variables; three (B, C, and F) in terms of ability, preference, and adjustment variables; and four (D, E, G, and H) in terms of preference and adjustment variables. This convergence of ability, preference, and adjustment measures lends further support to the thesis advanced in the articles describing the 1951 study—that these devices are measuring aspects of personality traits. More particularly, the association of scores from all three of these areas on three of the factors is again suggestive of a convergence of personality variables with those from the other two classes, such as was hinted at in the discussion of Factor F of the 1951 study (6, p. 197).

Table III shows the trait interrelations. Since the results of this study were intended for direct application, it was felt desirable to keep the computational formulae as simple as possible. Only for Trait B were more than four original variables used in the trait equations. The multiple regression coefficients between the traits and the factors ranged from

.68 to .88, the average coefficient being .77. These figures mean that the relations between the traits (expressed as correlation coefficients in Table III) are likely to move away from the relations between the factors (the cosines of the reference vectors of the final rotated oblique factor matrix). This did happen. Assuming that an orthogonal structure of trait relations represents an ideal, fifteen of the twenty-eight possible trait relations either retained the orthogonal relations of their parent factors, or moved in the direction of increased orthogonality (decreased relationship). Nine either retained the degree of interrelation shown by the equivalent factor cosines, or increased that relationship with no change in sign; and four, like their equivalent factors, showed substantial relations, but with a reversal in sign. It is difficult to evaluate the significance of these changes. Of the five largest, four involve Trait B (the pairs BD, BF, BG, BH), and two involve Trait D (the pairs BD, DC). Generally, however, the traits preserve a fairly high level of independence. For the total sample the trait coefficients range from .00 to .57 in magnitude, with nineteen of the twenty-eight being smaller than .19. The cosines for the rotated reference vectors ranged from .01 to .49 in magnitude, with thirteen of the twenty-eight being smaller than .19.

It is apparent from Table III that there can be considerable variation in trait interrelations between the various sub-groups. For example, the correlation for Traits B and D for the Science sample is .14, for the Commerce sub-group it is -.35. The correlation of Traits C and G is -.32 for the Arts group and .06 for the Commerce group. The correlation of Traits E and G is .03 for the Science sample and .52 for the Commerce sample. These differences suggest that sub-groups of a population may be identified or defined not only in terms of their positions on various measurable dimensions of behaviour but also in terms of characteristic interrelations between these dimensions.

Finally, it is to be noted that the study did produce useful equations for estimating academic achievement.

SUMMARY

Evidence was reviewed indicating that traits (defined as factorial composites of abilities and interests) showed significant relations to college achievement. The results of a new factorial study of measures of ability, interest, and adjustment were presented, and the complex factors which emerged were again interpreted as traits. It was shown that there were significant differences in trait scores between students registered in different faculties, and that there were significant relations between trait scores and achievement in the first year of college studies. Multiple regression equations were developed for estimating achievement in Arts, Science, and Commerce on the basis of trait scores.

REFERENCES

1. BENNETT, GEORGE K., SEASHORE, HAROLD G., & WESMAN, ALEXANDER G. *Differential aptitude tests*. New York: Psychological Corp., 1947.
2. BROWN, WILLIAM F., & HOLTZMAN, WAYNE H. *Survey of study habits and attitudes*. New York: Psychological Corp., 1956.
3. GORDON, LEONARD V. *Gordon personal profile*. Chicago: World Book Co., 1953.
4. KUDER, G. FREDERIC. *Kuder preference record: Vocational*. Chicago: Science Research Associates, 1942.
5. NELSON, M. J., & DENNY, E. C. *The Nelson-Denny reading test for colleges and senior high schools*. New York: Houghton Mifflin Co., 1929.
6. SMITH, D. D. Abilities and interests: I, A factorial study. *Canad. J. Psychol.*, 1958, 12, 191-201.
7. SMITH, D. D. Abilities and interests: II, Validation of factors. *Canad. J. Psychol.*, 1958, 12, 253-258.
8. THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.

SOME RUMINATIONS ON THE VALIDATION OF CLINICAL PROCEDURES¹

PAUL E. MEEHL

University of Minnesota

IT IS BECOMING ALMOST A CLICHÉ to say that "clinical psychology is in a state of ferment," a remark which is ambiguous as to whether the "ferment" is a healthy or pathological condition. Dr. E. Lowell Kelly finds upon follow-up that about 40 per cent of the young clinicians who were studied in the early days of the Veterans' Administration training programme now state that they would not go into clinical psychology if they had it to do over again (personal communication). In recent textbooks, such as Garfield's, one can detect a note of apology or defensiveness which was not apparent even a decade ago (13, pp. vi, 28, 88, 97, 101, 109, 116, 152, 166, 451, and *passim*). No doubt economic and sociological factors, having little to do with the substance of clinical psychology, contribute in some measure to this state of mind within the profession. But I believe that there are also deeper reasons, involving the perception by many clinicians of the sad state of the science and art which we are trying to practise (17). The main function of the clinical psychologist is psychodiagnosis; and the statistics indicate that, while the proportion of his time spent in this activity has tended to decrease in favour of therapy, it nevertheless continues to occupy the largest part of his working day. Psychodiagnosis was the original basis upon which the profession became accepted as ancillary to psychiatry, and it is still thought of in most quarters as our distinctive contribution to the handling of a patient. One is therefore disturbed to note the alacrity with which many psychologists move out of psychodiagnosis when it becomes feasible for them to do so. I want to suggest that this is only partly because of the even higher valence of competing activities, and that it springs also from an awareness, often vague and warded off, that our diagnostic instruments are not very powerful. In this paper I want to devote myself entirely to this problem, and specifically to problems of validity in the area broadly labeled "personality assessment."

I have chosen the word "ruminations" in my title. It helps from time to time for us to go back to the beginning and to formulate just what we

¹Invitational Address to the Canadian Psychological Association's Convention at Edmonton, Alberta, June 12, 1958.

are trying to do. I shall have to make some points which are perhaps obvious, but in the interest of logical completeness I trust that the reader will bear with me. In speaking about validity and validation, I shall employ the terminology proposed by the APA committee on test standards, making the fourfold distinction between predictive, concurrent, content, and construct validity. (1, see also 6.)

The practical uses of tests can be conveniently divided into three broad functions: *formal diagnosis* (the attachment of a nosological label); *prognosis* (including "spontaneous" recoverability, therapy-stayability, recidivism, response to therapy, indications for one kind of treatment rather than another); and *personality assessment* other than diagnosis or prognosis. This last function may be divided, somewhat arbitrarily, into *phenotypic* and *genotypic* characterization, the former referring to what we would ordinarily call the descriptive or surface features of the patient's behaviour, including his social impact; and the latter covering personality structure and dynamics, and basic parameters of a constitutional sort (for example, anxiety-threshold). Taking this classification of test functions as our framework, let us look at each one, asking the two questions: "Why do we want to know this?" and "How good are we at finding it out?"

Consider first the problem of formal psychiatric diagnosis. This is a matter upon which people often have strong feelings, and I should tell you at the outset that I have some prejudices. I consider that there are such things as disease entities in functional psychiatry, and I do not think that Kraepelin was as mistaken as some of my psychological contemporaries seem to think. It is my belief, for example, that there is a *disease*, schizophrenia, fundamentally of an organic nature, and probably of largely constitutional aetiology. I would explain the viability of the Kraepelinian nomenclature by the hypothesis that there is a considerable amount of truth contained in the system; and that, therefore, the practical implications associated with these labels are still sufficiently great, especially when compared with the predictive power of competing concepts, that even the most anti-nosological clinician finds himself worrying about whether a patient whom he has been treating as an obsessional character "is really a schizophrenic."

The fundamental argument for the utility of formal diagnosis can be put either causally or statistically, but it amounts to the same kind of thing one would say in defending formal diagnosis in organic medicine. One holds that there is a sufficient amount of aetiological and prognostic homogeneity among patients belonging to a given diagnostic group, so that the assignment of a patient to this group has probability implications which it is clinically unsound to ignore.

There are three commonly advanced objections to a nosological orientation in assessment, each of which is based upon an important bit of truth but which, as it appears to me, have been used in a somewhat careless fashion. It is first pointed out that there are studies indicating a low agreement among psychiatrists in the attachment of formal diagnostic labels. I do not find these studies very illuminating (2, 34, 38). If you are accustomed to asserting that "It is well known that formal psychiatric diagnoses are completely unreliable," I urge you to re-read these studies with a critical set as to whether they establish that thesis. The only study of the reliability of formal psychiatric diagnosis which approximates an adequate design is that of Schmidt and Fonda (48); and the results of this study are remarkably encouraging with regard to the reliability of psychiatric diagnosis. As these authors point out, some have inferred unreliability of formal diagnosis from unreliable assessment of other behavioural dimensions. Certainly our knowledge of this question is insufficient and much more research is needed.

I suppose that we are all likely to be more impressed by our personal experience than by what someone else reports when the published reports are not in good agreement and there is insufficient information to indicate precisely why they come to divergent results. For example, it is often said that the concept "psychopathic personality" is a waste-basket category that does not tell us anything about the patient. I know that many clinicians have used the category carelessly, and it is obvious that one who uses this term as an approximate equivalent to saying that the patient gets in trouble with the law is not doing anything very profound or useful by attaching a nosological label. I, on the other hand, consider the asocial psychopath (or, in the revised nomenclature, the sociopath) to be a very special breed of cat, readily recognized, and constituting only a small minority of all individuals who are in trouble because of what is socially defined as delinquent behaviour (in this connection see 31, 50). I consider it practically important to distinguish (a) a person who becomes legally delinquent because he is an "unlucky" sociopath, that is, got caught; (b) one who becomes delinquent because he is an acting-out neurotic; and (c) a psychiatrically normal person who learned the wrong cultural values from his family and neighbourhood environment.

Being interested in the sociopath, I have attempted to develop diagnostic skills in identifying this type of patient, and some years ago I ran a series on myself to check whether I was actually as good at it as I had begun to believe. I attempted to identify cases "at sight," that is, by observing their behaviour in walking down the hall or sitting in the hospital lounge, without conversing with the patient but snatching brief

samples of verbal behaviour and expressive movements, sometimes for a matter of a few seconds and never for more than five minutes. In the majority of cases I had no verbal behaviour at all. In the course of a year, I spotted 13 patients, as "psychopathic personality, asocial amoral type"; accepting staff diagnosis or an MMPI profile of psychopathic configuration as a disjunctive criterion, I was "correct" in 12 of the 13. This does not, of course, tell us anything about my false negative rate; but it does indicate that if I think a patient is a psychopath, there is reason to think I am correct. Now if I were interested in examining the "reliability" of the *concept* of the psychopathic personality, I should want to have clinicians like myself making the judgments.

Imagine, if you will, a psychologist trained to disbelieve in nosological categories and never alerted to those fascinating minor signs (lack of normal social fear, or what I call "animal grace," a certain intense, restless look about the eyes, or a score of other cues); suppose a study shows that such a psychologist tends not to agree with me, or that we both show low agreement with some second-year psychiatric resident whose experience with the concept has been limited to an hour lecture stressing the legal delinquency and "immaturity" (whatever that means) of the psychopath. What importance does such a finding have?

This matter of diagnostic skill involves a question of methodological presuppositions that is of crucial importance in interpreting studies of diagnostic agreement. The psychologist, with his tendency to an operational (20) or "pure intervening variable" type of analysis (32, 47) and from his long tradition of psychometric thinking in which reliability constrains validity, is tempted to infer directly from a finding that people disagree on a diagnostic label that a nosological entity has no objective reality. This is a philosophical mistake, and furthermore, it is one which would not conceivably be made by one trained in medical habits of thinking. When we move from the question of whether a certain sign or symptom should be given a high weight to the quite different question whether a certain disease entity has reality and is worth working hard to identify, disagreement between observers is (quite properly) conceived by physicians as *diagnostic error*. Neurological diagnoses by local physicians in outstate Minnesota are confirmed only approximately 75 per cent of the time by biopsy, exploratory surgery, or autopsy at the University of Minnesota Hospitals. The medical man does not infer from this result that the received system of neurological disease entities is unsound; rather he infers that physicians make diagnostic mistakes.

Furthermore, it is not even assumed that all of these mistakes could be eliminated by an improvement in diagnostic skill. One of the most highly skilled internists in Minneapolis (43) published a statistical

analysis of his own diagnoses over a period of 28 years based on patients who had come to autopsy. Imposing very stringent conditions upon himself (such as classifying a diagnostic error as eliminable if evidence could have been elicited by sufficient re-examination), he nevertheless found that 29 per cent of his diagnoses were errors which could not in principle have been eliminated because they fell in the category of "no evidence; symptoms or signs not obtained." How is this possible? Because not only are there diseases which are *difficult* to diagnose; there are individual cases which are for all practical purposes *impossible* to diagnose so long as our evidence is confined to the clinical and historical material.

Presumably anyone who takes psychiatric nosology seriously believes that schizophrenia (like paresis, or an early astrocytoma in a neurologically silent area) is an *inner state*, and that the correct attachment of a diagnostic label involves a probability transition from what we see on the outside to what is objectively present on the inside. The less that is known about the nature of a given disease, or the less emphasis a certain diagnostician gives to the identification of that disease, the more diagnostic errors we can expect will be made. That some psychiatrists are not very clever in spotting pseudoneurotic schizophrenia is no more evidence against the reality of this condition as a clinical entity than the fact that in 1850, long prior to the clinching demonstration of the luetic origin of paresis by Noguchi and Moore, even competent neurologists were commonly diagnosing other conditions, both functional and organic, as "general paralysis of the insane." By 1913 the luetic aetiology was widely accepted, and hence such facts as a history of chancre, secondary stage symptoms, positive spinal Wassermann, and the like were being given a high indicator weight in making the diagnosis (27). Yet the entity could not properly be *defined* by this (probable) aetiology; and those clinicians who remained still unconvinced were assigning no weight to the above-mentioned indicators. This must inevitably have led to diagnostic errors even by very able diagnosticians. It is impossible for diagnostic activity and research thinking to be suspended during the period—frequently long—that syndrome description constitutes our only direct knowledge of the disorder (33).

A second argument advanced against nosology is that it puts people in a *pigeon-hole*. I have never been able to understand this argument since whenever one uses *any* nomothetic language to characterize a human being one is, to that extent, putting him in a *pigeon-hole* (or locating him at a point in conceptual space); and, of course, every case of carcinoma of the liver is "unique" too. That some old-fashioned diagnosticians, untrained in psychodynamics, use diagnostic labels as a

substitute for understanding the patient is not an unknown occurrence, but what can one say in response to this except *abusus non tollit usum*? We cannot afford to decide about the merits of a conceptual scheme on the grounds that people use it wrongly.

A derivative of this argument is that diagnostic categories are not dynamics, and do not really tell us anything about what is wrong with the patient. There is some truth in this complaint, but again the same complaint could be advanced with regard to an organic disease concept at any stage in the development of the conception of it prior to the elucidation of its pathology and aetiology.

There is some confusion within our profession about the relation between content or dynamics and taxonomic categories. Many seem to think that when we elucidate the content, drives, and defences with which a patient is deeply involved, we have thereby explained why he is ill. But in what sense is this true? When we learn something about the inner life of a psychiatric patient, we find that he is concerned with aggression, sex, pride, dependence, and the like, that is, the familiar collection of human needs and fears. Schizophrenics are people, and if you are clever enough to find out what is going on inside a schizophrenic's head, you should not be surprised that these goings-on involve his self-image and his human relationships rather than, say, the weather. The demonstration that patients have psychodynamics, that they suffer with them, and that they deal with them ineffectively, does *not* necessarily tell us what is the matter with them, that is, why they are patients.

One is reminded in this connection of what happened when, after several years of clinicians busily over-interpreting "pathological" material in the TAT stories of schizophrenic patients, Dr. Leonard Eron took the pains to make a normative investigation and discovered that most of the features which had been so construed occurred equally or more often in a population of healthy college students (10).

There is no contradiction between classifying a patient as belonging to a certain taxonomic group and attempting concurrently to understand his motivations and his defences. Even if a certain major mental disease were found to be of organic or genetic origin, it would not be necessary to abandon any well-established psychodynamic interpretations. Let me give you an analogy. Suppose that there existed a colour-oriented culture in which a large part of social, economic, and sexual behaviour was dependent upon precise colour-discriminations. In such a culture, a child who makes errors in colour behaviour will be teased by his peer group, will be rejected by an over-anxious parent who cannot tolerate the idea of having produced an inferior or deviant child, and so on. One who was

unfortunate enough to inherit the gene for colour blindness might develop a colour neurosis. He might be found as an adult on the couch of a colour therapist, where he would produce a great deal of material which would be historically relevant and which would give us a picture of the particular pattern of his current colour dynamics. But none of this answers the question, "What is fundamentally the matter with these people?," that is, what do all such patients have in common? What they have in common, of course, is that defective gene on the X-chromosome; and this, while it does not provide a *sufficient* condition for a colour neurosis in such a culture, does provide the *necessary* condition. It is in this sense that a nosologist in that culture could legitimately argue that "colour neuroticism" is an inherited disease.

I think that none of these commonly heard objections is a scientifically valid reason for repudiating formal diagnosis, and that we must consider the value of the present diagnostic categories on their merits, on their relevance to the practical problems of clinical decision-making. One difficulty is that we do not have available for the validation of our instruments an analogue of the pathologist's report. It makes sense in organic medicine to say that the patient was actually suffering from disease X even though there was no evidence for it at the time of the clinical examination, so that the best clinician in the world could not have made a correct diagnosis on the data presented prior to autopsy. We have nothing in clinical psychology which bears close resemblance to the clinicopathological conference in organic medicine. Our closest analogue to pathology is "structure" and psychodynamics, and our closest analogue to the internist's concept of aetiology is a composite of constitution and learning history. If we had a satisfactory taxonomy of either constitution or learning history, we would be able to define what we meant by saying that a given patient is a schizophrenic. A well-established historical agent would suffice for this purpose, and Freud, for example, made an attempt at this in the early days (before he had realized how much of his patients' anamnesis was fantasy) by identifying the obsessional neurosis with a history of active and pleasurable erotic pre-pubescent activity, and hysteria with a history of passive and largely unpleasurable erotic experience (12).

Since anyone who takes formal diagnosis as a significant part of the psychologist's task must be thinking in terms of construct validity, (1, 6), he should have at least a vague sketch of the structure and aetiology of the disorders about which he speaks diagnostically. I do not think that it is appropriate to ask for an operational definition. My own view is that theoretical constructs are defined "implicitly" by the entire network of hypothesized laws concerning them; in the early stages of under-

standing a taxonomic concept, such as a disease, this network of laws is what we are trying to discover. Of course, when a clinician says, "I think this patient is really a latent schizophrenic," he should be able to give us some kind of picture of what he means by this statement. It could, however, be rather vague and still sufficient to justify itself at this stage of our knowledge. He might say:

I mean that the patient has inherited an organic structural anomaly of the proprioceptive integration system of his brain, and also a radical deficiency in the central reinforcement centres (or, to use Rado's language, a deficiency in his "hedonic capacity"). The combination of these proprioceptive and hedonic defects leads in turn to developmental disturbances in the body image and in social identification; the result at the psychological level being a pervasive disturbance in the cognitive functions of the ego. It is this defective ego-organization that is responsible for the primary associative disturbance set forth as the fundamental symptom of schizophrenia by Bleuler. The other symptoms of this disease, which may or may not be present, I would conceive as Bleuler does, and therefore my conception of the disorder is perhaps wider than is modal for American clinicians. By "pseudoneurotic schizophrenia" I would mean a patient with schizophrenia whose failure to demonstrate the accessory symptoms (and whose lower quantitative amount of even the primary symptoms) leads to his being readily misdiagnosed. Pseudoneurotic schizophrenia is just schizophrenia that is likely to go unrecognized.

Such a sketch is, to my mind, sufficient to justify the use of the schizophrenia concept at the present state of our knowledge. It is not very tight, and it is not intellectually satisfying. On the other hand, when combined with the set of indicators provided by Bleuler (3), Hoch and Polatin (21), and others, it is not much worse than the concept of general paresis as understood during most of the nineteenth century following Bayle's description in 1822. In this connection it is sometimes therapeutic for psychologists to familiarize themselves with the logicians' contributions to the methodological problems of so called "open concepts," "open texture," and "vagueness" (18, 19, 23, 41, 49, 57, 60). Even a slight acquaintance with the history of the more advanced sciences gives one a more realistic perspective on the relation of "operational" indicators to theoretical constructs during the early stages of a construct's evolution. (See, for example, 39, 45, 46, 56.)

The formal nosological label makes a claim about an inner structure or state; therefore, the concurrent validity of a test against our psychiatrist as criterion is not an end in itself, but rather is one piece in the pattern of evidence which is relevant to establishing the *construct* validity of both the test and the psychiatrist. If I really accept the psychiatric diagnosis as "*the criterion*," what am I doing with my test anyway? If I want to know what the psychiatrist is going to call patient Jones whom he has just finished interviewing, the obvious way to find

out is to leave my own little cubicle with its Rorschach and Multiphasic materials and walk down the hall to ask the psychiatrist what he is going to call the patient. This is a ludicrous way of portraying the enterprise, but the only thing which saves it from really being this way is that implicitly we reject concurrent validity with the psychiatrist's diagnosis as criterion, having instead some kind of construct validity in the back of our minds. The phrase "the criterion" is misleading. Because of the whole network of association surrounding the term "criterion," I would myself prefer to abandon it in such contexts, substituting the term "indicator." The impact of a patient upon a psychiatrist (or upon anyone else, for that matter) is one of a *family of indicators of unknown relative weights*; when we carry out a "validation" study on a new test, we are asking whether or not the test belongs to this family.

Note that the uncertainty of the link between nosology and symptom (or test) is a two-way affair. Knowing the formal diagnosis we cannot infer with certainty the presence of a given symptom or the result of a given test; conversely, given the result on a test, or the presence of a certain symptom, we cannot infer with certainty the nosology. (There are rare exceptions to this, such as thought-disorder occurring in the presence of an unclouded sensorium and without agitation, which I would myself consider pathognomonic of schizophrenia.) This uncertainty is found also in organic medicine, where there are very few pathognomonic symptoms and very few diseases which invariably show any given symptom. An extreme (but not unusual) example is the prevalence of those sub-clinical infections which are responsible for immunizing us as adults, but which were so "sub"-clinical that they were only manifested by a mild malaise and possibly a little fever, symptoms which, singly or jointly, do not enable us to identify one among literally hundreds of diagnostic possibilities.

One "statistical" advantage contributed by a taxonomy even when it is operating wholly at the descriptive or syndrome level is so obvious that it is easy to miss; I suspect that the viability of the traditional nosological rubrics, which could not be well defended upon aetiological grounds at present, is largely due to this contribution. When the indicators of membership in the class comprise a long list, none of which is either necessary or sufficient for the class membership, the descriptive information which is conveyed by the taxonomic name has a "statistical-disjunctive" character. That is, when we say that a patient belongs to category X, we are at least claiming that he displays indicators *a* or *b* or *c* with probability *p* (and separate probabilities *p_a*, *p_b*, and *p_c*). This may not seem very valuable, but considering how long it would take to convey to a second clinician the entire list of behaviour dispositions

whose probability of being present is materially altered by placing a patient in category X, we see that from the standpoint of sheer economy even a moderately good taxonomic system does something for us. More important in the long run is the fact that only a huge clinical team, with a tremendous amount of money to spend on a large number of patients over a long period of time, could hope to discover and confirm all $\frac{N(N-1)}{2}$ of the pair-wise correlations among the family of N indicators

that relate to the concept, to say nothing of the higher-order configural effects (22) that will arise in any such material. The research literature can yield cumulative knowledge and improvement of clinical practice in different settings by virtue of the fact that in one hospital an investigator, working with limited means, is able to show that patients diagnosed as schizophrenic tend to perform in a special way on a proverbs test; while another investigator in another hospital is showing that male patients diagnosed as schizophrenic have a high probability of reacting adversely to sexually attractive female therapists. Imagine a set of one hundred indicator variables and one hundred output variables; we would have to deal with ten thousand pair-wise correlations if we were to study these in one grand research project. The advantages in communicative economy and in cumulating research knowledge cannot, of course, be provided by a descriptive taxonomy which lacks intrinsic merit (that is, the syndrome does not objectively exist with even a moderate degree of internal tightness), or which, while intrinsically meritorious, is applied in an unskilful manner.

Let us turn now to our second main use of tests—prognosis. Sometimes the forecasting of future behaviour is valuable even if no special treatment is contemplated, because part of the responsibility of many clinical installations is to advise other agencies or persons, such as a court, as to the probabilities. But the main purpose of predictive statements is the assistance they give us in making decisions about how to treat a patient. Predictive statements of the form "If you treat the patient so-and-so, the odds are 8:2 that such-and-such will happen," will be with us for a very long time. As more knowledge about behavioural disorders is accumulated, we can expect a progressive refinement and differentiation of techniques; their differential impact will thereupon become greater, so that the seriousness of a mistake will be correspondingly increased. Furthermore, even if—as I consider highly unlikely but as we know some therapists are betting—it is discovered that for all patients the same kind of treatment is optimal, it is easily demonstrated from the statistics of mental illness, together with the most sanguine predictions as to the training of skilled professional personnel, that there will not be

adequate staff to provide even moderately intensive treatment for any but a minority of patients during the professional lifetime of anybody at present alive. So we can say with confidence that the decision to treat or not to treat will be a decision which clinicians are still going to be making when all of us have retired from the scene. As I read the published evidence, our forecasting abilities with current tests are not what you could call distinguished (see, for example, 61).

In connection with this problem of prognosis, let me hark back a moment to our discussion of formal nosology. One repeatedly hears clinicians state that they make prognostic decisions, not on the basis of a formal diagnosis but on their assessment of the individual's structure and dynamics. Where is the evidence that we can do this? So far as I am aware there is as much evidence indicating that one can predict the subsequent course of an illness from diagnostic categories (16) (or from crude life-history statistics) as there is that one can predict the course of an illness or the response to therapy from any of the psychological tests available. I should like to offer a challenge to any clinician who thinks that he can cite a consistent body of published evidence to the contrary.

In order to employ dynamic constructs to arrive at predictions, it would be necessary to meet two conditions. In the first place, we must have a sound theory about the determinative variables. Secondly, we must be in possession of an adequate technology for making measurements of those variables. As any undergraduate major in physics or chemistry knows, in order to predict the subsequent course of a physical system, it is necessary both to understand the laws which the system obeys and to have an accurate knowledge of the initial and boundary conditions of the system. Since clinical psychology is nowhere near meeting *either* of these two requirements, it must necessarily be poor at making predictions which are mediated by dynamic constructs. It is a dogma of our profession that we predict what people will do by understanding them individually, and this sounds so plausible and humanitarian that to be critical of it is like criticizing Mothers' Day. I can only reiterate that neither theoretical considerations nor the data available in the literature lend strong support to this idea in practice.

Let us turn to the third clinical task which the psychologist attempts to solve by the use of his tests, that of "personality assessment." Phenotypic characterization of a person includes the attribution of the ordinary clinical terms involving a minimal amount of inference, such as "patient hallucinates" or "patient has obsessional trends"; trait names from common English, such as the adjectives found in the lists published by Cattell (5, p. 219) or Gough (14); and, increasingly important in current

research, characterizations in the form of a single sentence or a short paragraph of the type employed by Stephenson (53), the Chicago Counseling Center (44), Block (4), and others. (Example: "The patient characteristically tries to stretch limits and see how much he can get away with.") A logical analysis of the nature of these phenotypic trait attributions is a formidable task although a very fascinating one. I am not entirely satisfied with any account which I have seen, or have been able to devise for myself. Perhaps not too much violence is done to the truth if we say that these are all in the nature of dispositional statements, the evidence for which consists of some kind of sampling, usually not representative, of a large and vaguely specified domain of episodes from the narrative that constitutes a person's life. It is complicated by the fact that even if we attempt to stay away from theoretical inferences, almost any single episode is susceptible of multiple classification under different families of atomic dispositions constituting a descriptive trait. The fact that the evidence for a trait attribution represents only a sample of the concrete episodes that exemplify atomic dispositions introduces an inferential element into such trait attributions, even though the trait name is intended to perform a purely summarizing rather than a theoretical function (6, pp. 292-3).

Phenotypic characterization presents a special problem which differentiates it from the functions of diagnosis and prognosis in the establishment of validity. Since it involves concurrent validity, its pragmatic justification is rather more obscure. Suppose we have a descriptive trait, say, "uncooperative with hospital personnel," an item which is not uncommon in various rating scales and clinical Q-pools in current use in the United States. Why administer an MMPI in order to guess, with imperfect confidence, whether or not the patient is being currently judged as uncooperative by the occupational therapist, the nursing supervisor, and the resident in charge of his case? This is even a more fruitless activity than our earlier example of using a test to guess the diagnosis given by the psychiatrist. From the theoretical point of view, the obvious reply is that the sampling of the domain of the patient's dispositions which is made by these staff members is likely to be deficient, both in regard to its *qualitative* diversity and representativeness as seen within the several contexts in which they interact with the patient, and *quantitatively* (simply from the statistical standpoint of size) during the initial portion of a patient's stay in the hospital. This reply leads to a suggestion concerning the design of studies which are concerned with phenotypic assessment from tests. Such designs should provide a "criterion" which is considerably superior in reliability to that which would routinely be available in the clinic on the basis of the

ordinary contacts. If it is concurrent validity in which we are really interested (upon closer examination this often turns out not to be the case), there is little point in administering a time-consuming test and applying the brains of a trained psychologist in order to predict the verbal behaviour of the psychiatric aid or the nurse. If it is our intention to develop and validate an instrument which will order or classify patients as to phenotypic features which are *not* reliably assessed by these persons in their ordinary contacts with the patient, then we need a design which will enable us to show that we have actually achieved this result.

As to the power of our tests in the phenotypic characterization of an individual, the available evidence is not very impressive when we put the practical question in terms of the *increment in valid and semantically clear information transmitted*. (See, for example, the studies by Kostlan (25), Dailey (8), Winch and More (58), Kelly and Fiske (24), Davenport (9), Sines (51), and Soskin (52).)

The question of concurrent validity in the phenotypic domain can be put at any one of four levels, in order of increasing practical importance. It is surprising to find that research on concurrent validity has been confined almost wholly to the first of these four levels. The weakest form of the validation question is, "How accurate are the semantically clear statements which can be reliably derived from the test?" It is a remarkable social phenomenon that we still do not know the answer to this question with respect to the most widely used clinical instruments. I do not see how anyone who examines his own clinical practice critically and who is acquainted with the research data could fail to make at least the admission that the power of our current techniques is seriously in doubt.

A somewhat more demanding question, which incorporates the preceding, would be: "To what extent does the test enable us to make, reliably, accurate statements which we cannot *concurrently* and *readily* (that is, at low effort and cost) obtain from clinical personnel routinely observing the patient *who will normally be doing so anyway* (that is, whose observations and judgments we will not administratively eliminate by the introduction of the test)?" In the preceding discussion regarding diagnosis and concurrent validity I oversimplified so grossly as to be a bit misleading. "How the staff rates" cannot be equated with "What the staff sees," which cannot in turn be equated with "What the patient does in the clinic"; and that, in turn, is not the equivalent of "What the patient does." If a patient beats his wife and does not tell his therapist about it, and the wife does not tell the social worker, the behaviour domain has been incompletely sampled by those making the ratings; they might *conclude* that he had beaten his wife, and this con-

clusion, while it is an inference, is still a conclusion regarding the phenotype. We cannot, of course, classify a certain concept as "theoretical" merely on the grounds that we have to make an inference in order to decide about a concrete instance of its application. This is a sampling problem, and therefore mainly (although not wholly) a matter of the time required to accumulate a sufficiently extensive sample. On the other hand, in our sampling of the patient's behavioural dispositions in the usual clinical context, it is not wholly a numerical deficiency in accumulation of episodes, because the sample which we obtain arises from a population of episodes that is in itself systematically biased. That is, the population of episodes which can be expected to come to our attention in the long run is itself a non-representative sub-population of all the behavioural events which constitute the complete narration of the patient's life.

A very stimulating paper is that of Kostlan (25). There are elements of artificiality in his procedure (of which he is fully aware) and these elements will no doubt be stressed by those clinicians who are determined to resist the introduction of adverse evidence. Nevertheless, his procedure was an ingenious compromise between the necessity of maintaining a close semblance to the actual clinical process, and a determination to quantify the incremental validity of tests. What he did, in a word, was to begin with a battery of data such as were routinely available in his own clinical setting and with which his clinicians were thoroughly familiar, consisting of a Rorschach, an MMPI, a sentence completion test, and a social case history. He then systematically varied the information available to his clinicians by eliminating one of these four sources at a time, arguing that the power of a device is probably studied better by showing the effect of its *subtraction* from the total mass of information than by studying it alone. The clinicians were required to make a judgment, from the sets of data presented to them, on each of 283 items which had been culled from a population of 1,000 statements found in the psychological reports written by this staff. The most striking finding was that on the basis of all three of these widely used tests his clinicians could make no more accurate inferences than they could make utilizing the Barnum effect (35, 8, 11, 52, 54, 55) when the all-important social history was deleted from their pool of data. A further fact, not stressed by Kostlan in his published report (but see 25 and 26), is that the absolute magnitude of incremental information, even when the results are statistically significant, is not impressive. For example, clinicians knowing only the age, marital status, occupation, education, and source of referral of a patient (that is, relying essentially upon Barnum effect for their ability to make correct statements) yield an average of about

63 per cent correct statements about the patient. If they have the Rorschach, Multiphasic, and Sentence Completion tests *but are deprived of the social case history*, this combined psychometric battery results in almost exactly the same percentage of correct judgments. On the other hand, if we consider their success in making inferences based on the social history together with the Sentence Completion test and the MMPI (that is, eliminating only the Rorschach, which made no contribution) we find them making 72 per cent correct inferences (my calculations from his Table 3), that is, a mere 9 per cent increment.

A thesis just completed at the University of Minnesota by Dr. Lloyd K. Sines is consistent with Kostlan's findings (51). Taking a Q-sort of the patient's therapist as his criterion, Sines investigated the contribution by a four-page biographical sheet, an MMPI profile, a Rorschach (administered by the clinician making the test-based judgments), and a diagnostic interview by this clinician. He determined the increment in Q-correlation with the criterion (therapist sort) when each of these four sources of information was inserted at different places in the sequence of progressively added information. The contribution of either of the two psychological tests, or both jointly, was small (and, in fact, knowledge of the Rorschach tended to exert an adverse effect upon the clinician's accuracy). For some patients, the application of a stereotype personality description based upon actuarial experience in this particular clinic provided a more accurate description of the patient than the clinician's judgment based upon any, or all, of the available tests, history, and interview data!

A third level of validation demand, in which we become really tough on ourselves, takes the form: "If there are kinds of clear non-trivial statements which can be reliably derived from the test, which are accurate, and which are not concurrently and readily obtainable by other means routinely available, *how much earlier in time* does the test enable us to make them?" It might be the case that we can make accurate statements from our tests at a time in the assessment sequence when equally trustworthy non-psychometric data have not accumulated sufficiently to make such judgments, but from the practical point of view there is still a need to know just how "advanced" this advance information is. So far as I know, there are no published investigations which deal with this question.

A final and most demanding way of putting the question, which is ultimately the practically significant one by which the contribution of our techniques must be judged, is the following: "If the test enables us to make reliably, clear, differentiating statements which are accurate and which we cannot readily make from routinely available clinical bases of

judgment; and if this additional information is not rapidly picked up from other sources during the course of continued clinical study of the patient; in what way, *and to what extent*, does this incremental advance information help us in treating the patient?" One might have a clear-cut positive answer to the first three questions and be seriously in error if he concluded therefrom that his tests were paying off in practice. On this fourth question, there is also no published empirical evidence.

In the absence of any data I would like to speculate briefly on this one. Suppose that a decision is made to undertake the intensive psychotherapy of a patient. A set of statements, either of a dichotomous variety or involving some kind of intensity dimension or probability-of-correctness, is available to the psychotherapist on the basis of psychological test results. How does the therapist make use of this knowledge? It is well known that competent therapists disagree markedly with regard to this matter, and plausible arguments on both sides have been presented. Presumably the value of such information will depend upon the kind of psychotherapy which is being practised; therapists of the Rogerian persuasion are inclined to believe that this kind of advanced knowledge is of no use; in fact they prefer to avoid exposure to it. Even in a more cognitively oriented or interpretative type of treatment, it may be argued that by the time the therapeutic interaction has brought forth sufficient material for interpretation and working-through to be of benefit to the patient, the amount of evidential support for a construction will be vastly greater than the therapist could reasonably expect to get from a psychological test report. It does not help the patient that there is "truth" regarding him in the therapist's head; since there is going to be a lot of time spent before the patient comes around to seeing it himself, and since this time will have to be spent regardless of what the therapist knows, perhaps there is no advantage in his knowing something by the second interview rather than by the seventh. On the other side, it may be argued that any type of therapy which involves even a moderate amount of selective attention and probing by the therapist does present moment-to-moment decision problems (for example, how hard to press, when to conclude that something is a blind alley, what leads to pick-up) so that advance information from psychometrics can set the therapist's switches and decrease the probability of making mistakes or wasting time. It seems to me that the armchair arguments pro and con in this respect are pretty evenly balanced, and we must await the outcome of empirical studies.

One rather disconcerting finding which I have recently come upon is the rapidity with which psychotherapists arrive at a stable perception of the patient which does not undergo much change as a result of subse-

quent contacts. I was interested in this matter of how early in the game the psychological test results enable us to say what the therapist *will be saying later on*. In our current research at Minnesota we are employing a Q-pool of 183 essentially "phenotypic" items drawn from a variety of sources. We are also using a "genotypic" pool of 113 items which consists of such material as the Murray needs, the major defence mechanisms, and various other kinds of structural-dynamic content. I was hoping to show that as the therapist learns more and more about his patient, his Q-correlation with the Q-description of the patient based upon blind analysis of the MMPI profile would steadily rise; furthermore, it is of interest to know whether there are *sub-domains* of this pool, such as mild and well-concealed paranoid trends, with respect to which the MMPI is highly sensitive early in the game. (From my own therapeutic work, I have the impression that a low Pa score has almost no value as an exclusion test, but that any patient, however non-psychotic he may be, who has a marked *elevation* on this scale will, sooner or later, present me with dramatic corroborating evidence.) However, I can see already that I have presented the test with an extraordinarily difficult task, because the Q-sorts of these therapists stabilize so rapidly. The therapists Q-described their patients after the first therapeutic hour, again after the second, then after the fourth, eighth, sixteenth, and twenty-fourth contact. If one plots the Q-correlation between each sorting and the sorting after twenty-four hours of treatment (or between each sorting and a pooled sorting; or between each sorting and the next successive sorting), one finds that by the end of the second or fourth hour, the coefficients with subsequent hours are pushing the sort-resort reliabilities. The convergence of the therapist's perception of his patient is somewhat faster in the phenotypic than in the genotypic pool, but even in the latter his conception of the patient's underlying structure, defence mechanisms, need-variable pattern, and so on seems to crystallize very rapidly. Even before examining the MMPI side of my data, I can say with considerable assurance that it will be impossible for the test to "prove" itself by getting ahead, and staying ahead, of the therapist to a significant extent. Of course, we are here accepting the psychotherapist's assessment as one which does converge to the objective truth about the patient in the long run, and this may not be true for all sub-domains of the Q-pool. The extent to which this rapid convergence to a stable perception represents invalid premature "freezing" is unknown (but see 7).

Personality characterization at the genotypic level will undoubtedly prove to be the most difficult test function to evaluate. A genotypic formulation, even when it is relatively inexplicit, seems to provide a kind of background which sets the therapist's switches as he listens to the

patient's discourse. What things he will be alert to notice, how he will construe them, what he will say and when, and even the manner in which he says it, are all presumably influenced by this complicated and partly unconscious set of perceptions and expectancies. Process research in psychotherapy is as yet in such a primitive state that one hardly knows even how to begin thinking about experiments which would inform us as to the pragmatic payoff of having advanced information, at various degrees of confidence, regarding specific features of the genotype. Even if it can be demonstrated that the therapist's perception of the patient tends with time to converge to that provided in advance by the test findings, this will never be more than a statistical convergence; therefore, in exchange for correctly raising the probability that one sub-set of statements is true of the patient, we will always be paying the price of expecting confirmation of some other unspecified sub-set which is erroneous.

Let me illustrate the problem by a grossly oversimplified example. Suppose that prior to either testing or interviewing, a dichotomously treated attribute has a base-rate probability of .60 in our particular clinic population. Suppose further that it requires an average of five therapeutic interviews before the therapist can reach a confidence of .80 with regard to the presence of this attribute. Suppose finally that a test battery yields this same confidence at the conclusion of diagnostic study (that is, before the therapy begins). During the five intervening hours, the therapist is presumably fluctuating in his assessment of this attribute between these two probability values, and his interview behaviour (as well as his inner cognitive processes) are being influenced by his knowledge of the test results. Perhaps because of this setting of his switches he is able to achieve a confidence around the .8 mark by the end of the fourth session, that is, two hours earlier than he would have been able to do without the test. Meanwhile, he has been concurrently proceeding in the same way with respect to a second attribute; but, unknown to him, in the present case the test is giving him misinformation about that attribute (which will happen in one patient out of five on our assumptions). It is impossible to say from our knowledge of the cognitive processes of interpretive psychotherapists, or from what we know of the impact of the therapeutic interaction upon the patient, whether a net gain in the efficacy of treatment will have been achieved thereby. The difficulties in unscrambling these intricate chains of cumulative, divergent (29), and interactive causation are enormous.

I suspect that the present status of process research in psychotherapy does not make this type of investigation feasible. Alternatively, we shift to "outcome" research. Abandoning an effort to understand the fine

causal details of the interaction between patient and therapist, we confine ourselves to the crude question, "Are the outcomes of psychotherapy influenced favourably, on the average, by making advance information from a psychometric assessment available to the therapist?" Granting the variability of patients and therapists, and the likely interaction between these two factors and the chosen therapeutic mode, it seems feasible to carry out factorial-design research in which this question might be answered with some degree of assurance. When so much of the clinical psychologist's time is expended in the effort to arrive at a psychodynamic formulation of the patient through the integration of psychological test data, to the point that in some out-patient settings the total number of hours spent on this activity is approximately equal to the median number of hours of subsequent therapeutic contact, I believe that we should undertake research of this kind without delay.

Whatever the future may bring with regard to the pragmatic utility of the genotypic information provided by psychometrics, I am inclined to agree with Jane Loevinger's view that tests should be constructed in a framework of a well-confirmed psychological theory and with attention devoted primarily to construct validity. In her recent monograph (28), Dr. Loevinger has suggested that it is inconsistent to lay stress on construct validity and meanwhile adopt the "blind, empirical, fact-to-fact" orientation I have expressed (35, 36). I do not feel that the cookbook approach is as incompatible with a dedication to long-term research aimed at construct validity as Dr. Loevinger believes. The future use of psychological tests, if they are to become more powerful than they are at present, demands, as Loevinger points out, cross-situational power. It would be economically wasteful to have clinicians in each of the hundreds of private and public clinical facilities deriving equations, actuarial tables, or descriptive cookbooks upon each of the various clinical populations. I would also agree with Loevinger that such cross-situational power is intimately tied to construct validity, and that the construction of a useful cookbook does not, in general, contribute appreciably to the development of a powerful theoretical science of chemistry.

On the other hand, there is room for legitimate disagreement, among those who share this basic construct-validity orientation, on an important interim question. If the development of construct-valid instruments which will perform with a high degree of invariance over different clinical populations hinges upon the elaboration of an adequate psychological theory concerning the domain of behaviour to be measured, then the rate of development of such instruments has a limit set upon it by the rate of development of our psychodynamic understanding. I per-

sonally am not impressed with the state of psychological theory in the personality domain, and I do not expect the edifice of personality constructs to be a very imposing one for a long time yet. Meanwhile, clinical time is being expended in the attempt to characterize patients by methods which make an inefficient use of even that modest amount of valid information with which our present psychometric techniques provide us.

The number of distinct attributes commonly viewed by clinicians as worth assessing is actually rather limited. The total number of distinguishable decision problems with which the psychiatric team is routinely confronted is remarkably small (see, for example, 8). It is not possible to say, upon present evidence, what are the practical limits upon the validity generalization of configural mathematical functions set up on large samples with respect to these decision classes. It is possible that the general *form* of such configural functions, and even the parameters, can be generalized over rather wide families of clinical populations, with each clinical administrator making correction of cutting scores or reassigning probabilities in the light of his local base-rates (37). One could tolerate a considerable amount of shrinkage in validity upon moving to a similar but non-identical clinical population without bringing the efficiency of an empirical cookbook down to the low level of efficiency manifested by clinicians who are attempting to arrive at such decisions on an impressionistic basis from the same body of psychometric and life history evidence. Halbower, for instance, showed that moving from an out-patient to an in-patient veteran population, while it resulted in considerable loss in the descriptive power of a cookbook based upon MMPI profile patterns, nevertheless maintained a statistically significant (and a practically important) edge over the Multiphasic reading powers even of clinicians who were working with the kind of population to which validity was being generalized (15). One of the things we ought to be trying is the joint utilization, in one function or table, of the most predictive kinds of life history data *together with* our tests. Some of the shrinkage in transition to allied but different clinical populations might be taken care of by the inclusion of a few rather simple and objective facts about the patient such as age, education, social class, referral source, percentage of service-connected disability, and the like.

Hence, I agree with Dr. Loevinger's emphasis upon the long-term importance of constructing tests which will be conceptually embedded in the network of psychological theory, and therefore superior in cross-situational power; in the meantime we do not have such tests, and there is some reason to think that in making daily clinical decisions a standard set of decision problems and trait attributions can be con-

structured. Such empirical research (readily within present limitations of personnel and theory) could result in the near future in cookbook methods which would include approximate stipulations as to those parametric modifications necessary for the main classes of clinical populations and for base rates, whether known or crudely estimated, in any given installation. I do not see anything statistically unfeasible about this, and I shall therefore continue to press for a serious prosecution of this line until somebody presents me with more convincing evidence than I have thus far seen that the clinical judge, or the team meeting, or the whole staff conference, is able somehow to surmount the limitations imposed by the inefficiency of the human mind in combining multiple variables in complex ways.

As for the long-term goal of developing construct-valid tests, maybe our ideas about the necessary research are insufficiently grandiose. Perhaps the kind of integrated psychometric-and-theory network which is being sought is not likely to be built up by the accumulation of a large number of minor studies. If we were trying to make a structured test scale, for instance, which would assess those aspects of a patient's phenomenology that are indicators of a fundamentally schizadaptive makeup, we would be carrying on an uphill fight against nature if we accepted as our criterion the rating of a second-year psychiatry resident on a seven-step "latent schizophrenia" variable! I would not myself be tempted to undertake the construction of an MMPI key for latent schizophrenic tendency unless I had the assurance that the classification or ordering of the patient population would be based upon a multiple attack taking account of all of the lines of evidence which would bear upon such an assessment in the light of my crude theory of the disease. *The desirability of a "criterion" considerably superior to what is routinely available clinically applies to the development of construct-valid genotypic measures even more than to criterion-oriented contexts.* Between such a hypothetical inner variable or state as "schizophrenic disposition," and almost any namable aspect of overt behaviour, there is interpolated quite a collection of nuisance variables. In order to come to a decision regarding, for example, a certain sub-set of cases which are apparently "test misses" (or which throw sub-sets of items in the wrong direction and hence provide evidence that those items should be modified or eliminated) one has to have a sufficiently good assessment of the relevant nuisance variables to satisfy himself that the apparent test or item miss is a miss in actuality.

This brings me to what I have often thought of as the curse of clinical psychology as a scientific enterprise. There are some kinds of psychological test construction or validation in which it suffices to know a very

little bit about each person, provided a large number of persons are involved (for example, in certain types of industrial, educational, or military screening contexts). At the other extreme, one thinks of the work of Freud, in which the most important process was the learning of a very great deal about a small number of individuals. When we come to the construction and validation of tests where, as is likely always to be true in clinical work, higher-order configurations of multi-variable instruments are involved, we need to know a great deal about each individual in order to come to a conclusion about what the test or item should show regarding his genotype. However, in order to get statistical stability for our weights and to establish the reality of complex patterning trends suggested by our data, we need to have a sizable sample of individuals under study. So that where some kinds of psychological work require us to know only a little bit about a large number of persons, and other kinds of work require us to know a very great deal about a few persons, construct validation of tests of the sort that Loevinger is talking about will probably require that we know a great deal, and at a fairly intensive or "dynamic" level, about a large number of persons. You will note that this is not a reflection of some defect of our methods or lack of zeal in their application but arises, so to speak, from the nature of things. I do not myself see any easy solution to this problem.

I am sure that by now you are convinced of the complete appropriateness of my title. I am aware that the over-all tenor of my remarks could be described as somewhat on the discouraged side. But we believe in psychotherapy that one of the phases through which most patients have to pass is the painful one between the working through of pathogenic defences and the reconstitution of the self-image upon a more insightful basis. The clinical psychologist should remind himself that medical diagnostic techniques frequently have only a modest degree of reliability and validity. I have, for instance, recently read a paper written by three nationally known roentgenologists on the descriptive classification of pulmonary shadows, which these authors subtitle "A Revelation of Unreliability in the Roentgenographic Diagnosis of Tuberculosis" (40). I must say that my morale was improved after reading this article.

In an effort to conclude these ruminations on a more encouraging note, let me try to pull together some positive suggestions. Briefly and dogmatically stated, my constructive proposals would include the following:

1. Rather than decrying nosology, we should become clinical masters of it, recognizing that some of our psychiatric colleagues have in recent times become careless and even unskilled in the art of formal diagnosis.

2. The quantitative methods of the psychologist should be applied to

the refinement of taxonomy and not confined to data arising from psychological tests. (I would see the work of Wittenborn (59) and of Lorr and his associates (30) as notable beginnings in this direction.)

3. While its historical development typically begins with syndrome description, the reality of a diagnostic concept lies in its correspondence to an inner state, of which the symptoms or test scores are fallible indicators. Therefore, the validation of tests as diagnostic tools involves the psychiatrist's diagnosis merely as one of an indicator family, not as a "criterion" in the concurrent validity sense. Accumulation of numerous concurrent validity studies with inexplicably variable hit-rates is a waste of research time.

4. Multiple indicators, gathered under optimal conditions and treated by configural methods, must be utilized before one can decide whether to treat inter-observer disagreement as showing the unreality of a taxonomy or merely as diagnostic error.

5. We must free ourselves from the almost universal assumption that when we elucidate the motives and defences of a psychiatric patient, we have thereby explained why he has fallen ill. As training analysts have observed for years, patients and "normals" tend to have pretty much the same things on their minds, conscious and unconscious.

6. The relative power, for prognosis and treatment selection, of formal diagnosis, non-nosological taxonomies based upon trait clusters, objective life-history factors, and dynamic understanding via tests, is an empirical question in need of study, rather than a closed issue. We must face honestly the disparity between current clinical practice and what the research evidence shows about the relatively feeble predictive power of present testing methods.

7. There is some reason to believe that quantitative treatment of life-history data may be as predictive as psychometrics in their present state of development. Research along these lines should be vigorously prosecuted.

8. It is also possible that interview-based judgments at a minimally inferential level, if recorded in standard form (for example, Q-sort) and treated statistically, can be made more powerful than such data treated impressionistically as is currently the practice.

9. While maximum generalizability over populations hinges upon high construct validity in which the test's functioning is imbedded in the network of personality theory, there is a pressing interim need for empirically derived rules for making clinical decisions (that is, "clinical cookbooks"). Research is needed to determine the extent to which such cookbooks are tied to specific clinic populations and how the recipes can be adjusted in moving from one population to another.

10. Perhaps there are mathematical models, more suitable than the factor-analytic one and its derivatives, for making genotypic inferences, and especially inferences to nosology. Investigation of such possibilities must be pursued by psychologists who possess a thorough familiarity with the intellectual traditions of medical thinking, a solid grasp of psychodynamics, and enough mathematical skill to take creative steps along these lines.

11. From the viewpoint of both patients' welfare and taxpayers' economics, the most pressing *immediate* clinical research problem is that of determining the incremental information provided by currently used tests, especially those which consume the time of highly skilled personnel. We need not merely validity, but incremental validity; further, the temporal factor, "Does the test tell us something we are not likely to learn fairly early in the course of treatment?" should be investigated; finally, it is well within the capacity of available research methods and clinical facilities to determine what, if any, is the pragmatic advantage of a personality assessment being known in advance by the therapist.

12. In pursuing these investigations we might better avoid too much advertising of the results since neither psychiatrists nor government officials are in the habit of evaluating the efficiency of their own procedures, a fact which puts psychologists at a great propaganda disadvantage while the science is still in a primitive stage of development.

REFERENCES

1. APA COMMITTEE ON TEST STANDARDS. Technical recommendations for psychological tests and diagnostic techniques. *Psychol. Bull. Suppl.*, 1954, 51, 2, Part 2, 1-38.
2. ASH, P. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1949, 44, 272-276.
3. BLEULER, E. *Dementia praecox*. New York: International Univer. Press, 1950.
4. BLOCK, J., & BAILEY, D. *Q-sort item analyses of a number of MMPI scales*. Technical Memorandum OERL-TM-55-7 Officer Education Research Laboratory. Air Force Personnel and Training Research Center, Air Research and Development Command, Maxwell Air Force Base, Alabama, 1955.
5. CATTELL, R. B. *Description and measurement of personality*. New York: World Book Company, 1946.
6. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, 52, 281-302.
7. DAILEY, C. A. The effect of premature conclusion upon the acquisition of understanding a person. *J. Psychol.*, 1952, 33, 133-152.
8. DAILEY, C. A. The practical utility of the clinical report. *J. consult. Psychol.*, 1953, 17, 297-302.
9. DAVENPORT, BEVERLY F. The semantic validity of TAT interpretations. *J. consult. Psychol.*, 1952, 16, 171-175.

10. ERON, L. D. Frequencies of themes and identifications in the stories of schizophrenic patients and non-hospitalized college students. *J. consult. Psychol.*, 1948, 12, 387-395.
11. FORER, B. R. The fallacy of personal validation: A classroom demonstration of gullibility. *J. abnorm. soc. Psychol.*, 1949, 44, 118-123.
12. FREUD, S. Further remarks on the defense neuro-psychoses. *Collected papers*, I, 155-182. London: Hogarth Press, 1948.
13. GARFIELD, S. *Introductory clinical psychology*. New York: Macmillan, 1957.
14. GOUGH, H. G., MCKEE, M. G., & YANDELL, R. J. *Adjective check list analyses of a number of selected psychometric and assessment variables*. Institute of Personality Assessment and Research. Berkeley: Univer. California, 1953.
15. HALBOWER, C. C. A comparison of actuarial versus clinical prediction to classes discriminated by MMPI. Unpublished Ph.D. thesis, Univer. Minnesota, 1955.
16. HASTINGS, D. W. Follow-up results in psychiatric illness. *Amer. J. Psychiat.*, 1958, 114, 1057-1066.
17. HATHAWAY, S. R. A study of human behavior: the clinical psychologist. *Amer. Psychologist*, 1958, 13, 257-285.
18. HEMPEL, C. G. Problems and changes in the empiricist criterion of meaning. *Revue internat. philosophie*, 1950, 4, 41-63.
19. HEMPEL, C. G. Fundamentals of concept formation in empirical science. *International encyclopedia of unified science*, II, no. 7. Chicago: Univer. Chicago Press, 1952.
20. HEMPEL, C. G. A logical appraisal of operationism. *Scientific Mon.*, 1954, 79, 215-220.
21. HOCH, P. & POLATIN. Pseudoneurotic forms of schizophrenia. *Psychiat. Quart.*, 1949, 23, 248-276.
22. HORST, P. Pattern analysis and configural scoring. *J. clin. Psychol.*, 1954, 10-11.
23. KAPLAN, A. Definition and specification of meaning. *J. Philosoph.*, 1946, 43, 281-288.
24. KELLY, E. L. & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor, Mich.: Univer. Michigan Press, 1951.
25. KOSTLAN, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, 18, 83-88.
26. KOSTLAN, A. A reply to Patterson. *J. consult. Psychol.*, 1955, 19, 486.
27. KRAEPELIN, E. *General paresis* (Trans. J. W. MOORE). New York: Nervous and Mental Disease Publishing Co., 1913.
28. LOEVINGER, JANE. Objective tests as instruments of psychological theory. *Psychol. Reports, Monogr. Suppl.* 9, 1957, 3, 635-694.
29. LONDON, I. D. Some consequences for history and psychology of Langmuir's concept of convergence and divergence of phenomena. *Psychol. Rev.*, 1946, 53, 170-188.
30. LORR, M. & RUBINSTEIN, E. A. Factors descriptive of psychiatric outpatients. *J. abnorm. soc. Psychol.*, 1955, 51, 514-522.
31. LYKKEEN, D. T. A study of anxiety in the sociopathic personality. *J. abnorm. soc. Psychol.*, 1957, 55, 6-10.
32. MACCORQUODALE, K. & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.
33. MAJOR, R. H. *Classic descriptions of disease*. Springfield, Ill.: Charles C. Thomas, 1932.

34. MASSERMAN, J. H. & CARMICHAEL, H. T. Diagnosis and prognosis in psychiatry with a follow-up study of the results of short-term general hospital therapy of psychiatric cases. *J. ment. Sci.*, 1939, 84, 893-946.
35. MEEHL, P. E. Wanted—a good cookbook. *Amer. Psychologist*, 1956, 11, 263-272.
36. MEEHL, P. E. When should we use our heads instead of the formula? *J. consult. Psychol.*, 1957, 4, 268-273.
37. MEEHL, P. E. & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.*, 1955, 52, 194-216.
38. MEHLMAN, B. The reliability of psychiatric diagnosis. *J. abnorm. soc. Psychol.*, 1952, 47, 577-578.
39. NASH, L. K. *The atomic-molecular theory*. Cambridge: Harvard Univer. Press, 1950.
40. NEWELL, R. R., CHAMBERLAIN, W. E., & RIGLER, C. Descriptive classification of pulmonary shadows: a revelation of unreliability. *Amer. Rev. Tuberculosis*, 1954, 69, 566-584.
41. PAP, A. Reduction sentences and open concepts. *Methodos*, 1953, 5, 3-30.
42. PATTERSON, C. H. Diagnostic accuracy or diagnostic stereotype? *J. consult. Psychol.*, 1955, 19, 483-485.
43. PEPPARD, T. A. Mistakes in diagnosis. *Minnesota Med.*, 1949, 32, 510-11.
44. ROGERS, C. R. & DYMOND, R. F. *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
45. ROLLER, D. E. *The development of the concept of electric charge*. Cambridge: Harvard Univer. Press, 1954.
46. ROLLER, D. E. *The early development of the concepts of temperature and heat*. Cambridge: Harvard Univer. Press, 1950.
47. ROZEBOOM, W. Mediation variables in scientific theory. *Psychol. Rev.*, 1956, 63, 249-264.
48. SCHMIDT, H. O. & FONDA, C. P. Reliability of psychiatric diagnosis: A new look. *J. abnorm. soc. Psychol.*, 1956, 52, 262-267.
49. SCRIVEN, M. Definitions, explanations, and theories. In H. FEIGL, M. SCRIVEN, & G. MAXWELL, *Concepts, theories and the mind-body problem*. Minnesota Studies in the Philosophy of Science, II. Minneapolis: Univer. Minnesota Press, 1958, pp. 99-195.
50. SIMONS, D. J. & DIETHELM, O. Electroencephalographic studies of psychopathic personalities. *Arch. Neurol. & Psychiat.*, 1946, 55, 619-627.
51. SINES, L. K. An experimental investigation of the relative contribution to clinical diagnosis and personality description of various kinds of pertinent data. Unpublished Ph.D. thesis, Univer. Minnesota, 1957.
52. SOSKIN, W. F. Bias in past-diction from projective tests. *J. abnorm. soc. Psychol.*, 1954, 49, 69-74.
53. STEPHENSON, W. The significance of Q-technique for the study of personality. In M. L. REYMERT (ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.
54. SUNDBERG, N. The acceptability of fake versus bona fide personality test interpretations. *J. abnorm. soc. Psychol.*, 1955, 50, 145-147.
55. TALLENT, N. On individualizing the psychologist's clinical evaluation. *J. clin. Psychol.*, 1958, 14, 243-44.
56. TAYLOR, L. W. *Physics, the pioneer science*. New York: Houghton Mifflin Co., 1941.
57. WAISMANN, F. Verifiability. *Proc. Aristotelian Soc. Suppl.*, 1945, 19, 119-150.

58. WINCH, R. F. & MORE, D. M. Does TAT add information to interviews? Statistical analysis of the increment. *J. clin. Psychol.*, 1956, 12, 316-321.
59. WITTENBORN, J. R. *Wittenborn Psychiatric Rating Scales*. New York: Psychological Corp., 1955.
60. WITTGENSTEIN, L. *Philosophical investigations*. Oxford: Blackwell, 1953.
61. ZUBIN, J. & WINDLE, C. Psychological prognosis of outcome in the mental disorders. *J. abnorm. soc. Psychol.*, 1954, 49, 272-281.

BOOK REVIEWS

Perception and Communication. By D. E. BROADBENT. New York: Pergamon Press, 1957. Pp. v, 338. \$8.50.

D. E. BROADBENT, employed by the British Medical Research Council as a research psychologist for eight or nine years, and recently appointed Director of its Applied Psychology Research Unit, has, as the writer of some forty publications, contributed considerably to our knowledge of man's behaviour. His book, *Perception and Communication*, summarizes his experimental findings in one major field of inquiry: attention, in general; the study of multiple stimulation, in particular. His justification for this specialization is nicely contained in an apt quotation from Sherrington, "the interference of unlike reflexes and the alliance of like reflexes in their action upon their common paths seem to be at the very root of the great psychical forces of attention."

The author also attempts to relate his own findings and conclusions to the work of others in the same area, and in contiguous areas. Finally, he considers the theoretical implications of these studies.

These aims and achievements delineate the value of the book, a happy coalescence of the practical and the theoretical. We are given a fund of information concerning the process of attending. This information is intrinsically useful and interesting, and of itself would offer sufficient justification both for writing the book, and for its careful perusal by all psychologists. But over and above this we are offered insights into a more general theory of behaviour, for the pervasive nature of attention is made obvious. The author's explorations of these theoretical implications lead him into several areas as a glance at the chapter headings shows. "The General Nature of Vigilance," "The Nature of Extinction," "Immediate Memory and the Shifting of Attention," "The Selective Nature of Learning," "Recent Views on Skill," to name a few. Nor is the treatment of these topics superficial. It offers enough breadth of review and depth of analysis as to be almost definitive.

In this respect the choice of title was perhaps unfortunate. One feels that the specializing psychologist's perception of the author's communication may not assure the book the wide audience it undoubtedly deserves.

P. J. FOLEY

Defence Research Medical Laboratories
Toronto, Ontario

Thinking: An Experimental and Social Study. By SIR FREDERIC BARTLETT.
London: Allen & Unwin Ltd., 1958. Pp. 11, 203. 18s.

CONTEMPORARY PSYCHOLOGICAL THEORY may be said to gain higher status the more it makes use of variables which have "tight" operational clarity. Professor Bartlett chooses to study thinking with "loose" definitions of variables and to restrict his variables to the dependent side almost exclusively. His book is a descriptive study of the thinking process using few mediational postulates and no guesses as to how thinking works neuro-physiologically. I think it would miss the point of the study to seriously criticize the descriptive approach or the facts that (1) what Bartlett calls "experiments" are simple demonstrations, (2) no care is given to sampling of subjects or probability statements in the interpretation of findings, (3) distant analogies give the study its orientations, etc. It is wiser to assume that Bartlett is aware of this supposed status hierarchy in modern psychology and has *chosen* to ignore it. Then the value of the work is seen to reside in its broadness. It is an over-view of the problem coming from a wise and outstanding psychologist who has obviously thought long and hard about thought.

Bartlett argues that the best approach to the study of the thinking processes is to use as a guide some well-studied simpler, but related, form of behaviour. This procedure saves us from the construction of some "general theory or other" which is likely to be tautological and arbitrary. Thinking is then considered as a complex ability or skill and various properties of skilled behaviour (such as "timing," "stationary phases," "the point of no return," and "direction") are distinguished so that they may be demonstrated in the thinking process. In the introductory chapter, a theoretical chapter and in a few pages of final remarks, this analogy is pursued and Bartlett concludes that "it has seemed not only that thinking of all kinds possesses (these properties of skill), but also that their study does throw some real light upon the thinking processes themselves" (p. 198). This reviewer was not so impressed by the similarity of the two processes or by this approach to thought. It would be unfair, however, not to mention that the possibilities for greater knowledge about thinking may very well come from further research and theory about abilities and skill. In fact the studies of Vince, Mackworth, and Poulton (especially the concept of "perceptual anticipation") appear to have broad theoretical significance which may well throw light on thinking processes.

It strikes me that it is when Professor Bartlett presents his broader view of thinking that he throws light on the subject. He distinguishes between "thinking within closed systems," which includes interpolation or gap-filling, extrapolation, and making use of evidence in disguise, and

"adventurous thinking" which samples the thinking of the scientist, the artist, and the "everyday" thinker. Consider the following puzzle: two 6-place "numbers" are to be added together, but letters are used instead of digits, e.g., DONALD + GERALD = ROBERT. One letter only is assigned a number but some letters are repeated. You are given the clue that $D = 5$. As one works this out he tries all sorts of seemingly personal approaches, but Bartlett's demonstrations using a number of subjects suggest that there is a predictable pattern of approaches. Comparable patterns are noted in the solution of sectional map-reading problems. Even scientific thinking appears to fall into similar patterns of approach. In his chapter on scientific thinking Bartlett traces the "approaches" of groups of scientists studying bacteria in one case and reaction time in the other.

His section on the characteristics of experimental thinking is brilliant, and his discussion of thinking "with a social theme" will certainly suggest new directions of thought for social and perceptual psychologists.

It is my prediction that this book will have an important influence on psychology, not immediately but in time. It will take time to make use of the insights presented in view of the current trend toward operational tightness. I also wonder how many full professors in the North American scene are actually, themselves, experimenting with real, live subjects as Bartlett obviously is.

WALLACE E. LAMBERT

McGill University

Principles of Perception. By S. HOWARD BARTLEY. New York: Harper and Brothers, 1958. Pp. xii, 482. \$6.50.

PSYCHOLOGISTS TEACHING COURSES in perception and in experimental psychology have waited for years for a textbook which would cover this broad field which extends from quantum theory to social psychology. Unfortunately, we must still wait.

In *Principles of Perception* Bartley begins with the avowal that perception is to be treated as a part of human behaviour, a biological science. This broad definition permits him to introduce the complex stimuli of social perception. This is a welcome innovation, and a major contribution. Attention is also given to a number of traditional topics: the idea of threshold, the definition of stimulus, and the problem of isolating perception from other aspects of behaviour. The student is urged to clarify his thinking by using equivalent but distinct terms for the physical and the phenomenal worlds, for example, intensity and brightness, pulse and flash.

The early chapters cover many topics: theories of perception (F. Allport's divisions are rather closely followed here), problems of epistemology (unfortunately without reference to the laws of specific nerve energies), signs and symbols as objects of perception, and sensory interaction. Chapter v contains many experiments demonstrating the importance of learning in the development of perception.

The next twelve chapters—two-thirds of the book—cover the traditional areas of perception in the traditional manner, the one exception to this being the omission of the basic psychophysical methods and laws. The Weber fraction is exiled to a chapter on muscular mechanisms. (This practically precludes the use of the book in an experimental course.) Otherwise, there is much solid stuff: photometry and radiometry, brightness discrimination, constancies, hearing, vision, and the other senses. The chapters on visual acuity and space perception are excellent. The chapters on hearing and colour vision suffer from a failure to relate the phenomenal data to underlying sensory and physiological mechanisms. It is the relation between facts, not the facts themselves, that makes science the fascinating endeavour that it is.

In the final eighty pages, Bartley returns to the broader problems of perception: social perception, individual differences in perception, and anomalies of perception. Some of this material is accepted uncritically: for example, the conclusion from the Lazarus and McCleary experiment which supposedly demonstrated the phenomenon of "subception" is allowed to stand without mention of the excellent experimental refutation by Bricker and Chapanis.

Unfortunately, the book has little to recommend it in respect to communication, mostly because the chapters vary widely in level of difficulty. A few chapters are excellent for seniors; others are too elementary, too repetitious, and too patronizing for anyone beyond an introductory course. Students are exasperated, not educated, by a passage (and there are a number of them) such as this: "In Fig. 6.7, for example, it will be noted that the labeling of the two axes is given in symbol form. One of the labels is an I. This stands for intensity. . . ." Surely elementary politeness dictates that the student be paid more respect than this.

The prime requisite of a textbook is that it reflect the ideas and facts in a given area. This the book fails to do. Perceptionists seek the answers to two questions: what is a stimulus that a man may know it?; What is a man that he may know a stimulus? To answer these questions, perceptionists are turning to the phenomenology of complex perception, the physiology of the sensory system, and the mathematics of communication theory and sensory scaling. No hint of these activities appears in this book.

No mention is made of the studies by Michotte on the perception of causality, or of the research on perceptual alertness and vigilance. Scant attention is paid the physiologist. (This failure is surprising, for Bartley, himself, did much of the pioneer work which mated neurophysiology and psychophysics.) Granit, Hartline, and others have opened a new field. Old problems in perception, such as hearing, colour vision, successive and simultaneous contrast, visual acuity, and even alertness and attention, are yielding their secrets to the psychophysiologist armed with his micro-electrode and oscilloscope. None of this appears in the book.

On the mathematical side of perception, no mention is made of the successful attempt by Attneave to quantify Gestalt concepts in terms of information theory. Also nothing appears on the successful application of the signal-to-noise-ratio model in hearing and vision. In spite of the lengthy passages devoted to complex stimuli, Bartley fails to consider what measurement techniques are feasible in this area. The future of measurement of social perception lies in the scaling techniques being developed by Coombs and others.

In sum, the book does not adequately cover its subject, namely, perception. If the book is to be used, it seems most suited for a second course for those going no further in psychology—and even then much material will have to be added by the instructor.

W. CRAWFORD CLARK

University of Michigan

Family Relationships and Delinquent Behavior. By F. IVAN NYE. New York: John Wiley and Sons, Inc., 1958. Pp. xii, 168. \$4.95.

PENDANT TRÈS LONGTEMPS, la psychopathologie s'est confinée entre les murs des asiles d'aliénés et n'a fait porter ses observations que sur les cas extrêmes que l'on y rencontre. Elle ne s'intéressait qu'à un groupe restreint, négligeant de considérer toute la gamme des comportements irrationnels ou symptomatiques, tels qu'ils se manifestent dans les cadres de la vie quotidienne. C'est un fait que la même méthode d'approche tend encore à prévaloir dans notre étude de la conduite antisociale qui, le plus souvent, ne tient compte que du repris de justice ou du jeune délinquant des écoles de protection. Nye et ses collaborateurs ont donc fait un effort éminemment utile en s'affranchissant de ces visières, qui nous empêchent de mesurer la véritable ampleur d'un phénomène inquiétant. Avec l'impitoyable rigueur de leurs tabulations statistiques, ils ont levé le voile sur les désordres de conduite que l'on peut s'attendre

à rencontrer chez la masse des adolescents qui fréquentent une école ordinaire.

Recourant ensuite aux procédés de quantification les plus minutieux et les plus raffinés, ils ont fait ressortir le caractère nettement relatif et provisoire de certaines conclusions auxquelles avaient abouti les recherches antérieures et que l'on avait pris l'habitude de regarder comme définitivement acquises. Nous songeons particulièrement, ici, aux répercussions malheureuses généralement attribuées aux carences socio-économiques ou à celle de certaines structures familiales. Dans la même ligne, l'effort consciencieux qui a été tenté pour mieux apprécier l'influence des divers types de relations parent-enfant sur l'orientation de la conduite, au cours de l'adolescence, mérite sûrement toute notre admiration. Il permet de se rendre compte, une fois de plus, qu'au plan social et dans la mesure où elle est provoquée par un concours de circonstances extrinsèques à l'individu lui-même, la délinquance est un phénomène beaucoup trop complexe pour être circonscrit à l'aide de mesures unilatérales. L'ensemble des données d'observation recueillies pointent dans le sens d'une conclusion dont on ne saurait exagérer l'importance : une prévention efficace de la délinquance juvénile repose sur un système multiforme de contrôle, direct ou indirect, des activités de nos adolescents et exige un effort concerté de tous les organismes concernés.

Si utile et si éclairante que puisse paraître cette enquête sociologique, il convient toutefois de rappeler qu'elle n'apporte aucune explication quant aux aspects essentiels du problème. La délinquance ne saurait être envisagée exclusivement comme une variable dont l'allure peut être étudiée en fonction de toute une diversité de facteurs présentant une corrélation plus ou moins marquée avec elle. Comme le symptôme, elle se présente comme la manifestation d'attitudes ou de fonctionnements de la personnalité, susceptibles d'être regardés tantôt comme réversibles, tantôt comme irréversibles. C'est dire que la conduite délinquante que l'on parviendra à influencer par des mesures de contrôle extrinsèques, si répandue soit-elle, reste vraisemblablement dans les limites d'une irrationalité qui persiste normalement en chacun de nous et sur laquelle la moralité établit progressivement son emprise. Quant à la délinquance, au sens strict du terme et en tant qu'elle désigne un type caractéristique de comportement humain, elle ne peut être considérée comme un phénomène sociologique que par répercussion. Elle nous replace devant un problème essentiellement individuel que le psychologue devra étudier en profondeur : celui de la genèse, des vicissitudes et de la dynamique de notre socialisation.

NOËL MAILLOUX, O.P.

Université de Montréal

ux
et
re-
me
er-
co-
me
ier
de
tre
an
de
un
de
ies
er
se
és
es

ie,
on
ait
re
ne
le
ts
es,
ue
si
a-
la
e,
ne
é-
an
er
le
p.